

Scenery entropy as an invariant of RWRS processes

Tim Austin*

Courant Institute, New York University,
New York, NY 10012, U.S.A.
tim@cims.nyu.edu

Abstract

Probabilistic models of random walks in random sceneries give rise to examples of probability-preserving dynamical systems. A point in the state spaces consists of a walk-trajectory and a scenery, and its ‘motion’ corresponds to shifting the time-origin.

These models were proposed as natural examples of non-Bernoulli K-automorphisms by Adler, Ornstein and Weiss. This was proved in a famous analysis by Kalikow using Ornstein’s Very Weak Bernoulli characterization of Bernoulli processes. Since then, various authors have generalized this construction to give other examples, including some smooth examples due to Katok and Rudolph.

However, the methods used to prove non-Bernoullicity do not obviously show that these examples are distinct from one another. This paper introduces a new isomorphism-invariant of probability-preserving systems, and shows that in a large class of the above examples it essentially captures the Kolmogorov-Sinai entropy of the scenery process alone. As a result, constructions that use different scenery-entropies give continuum-many non-isomorphic examples. Conditionally on an invariance principle for certain local times, these include a continuum of distinct smooth non-Bernoulli K-automorphisms on a fixed compact manifold.

Contents

1	Introduction	2
2	Preliminaries: analysis and probability	11

*Research supported by a fellowship from the Clay Mathematics Institute

3 Preliminaries: ergodic theory	17
4 Informal discussion of the RWRS marginal metrics	29
5 The new invariant	35
6 The combinatorial basis of the upper bound	49
7 The upper bound	56
8 Meandering of cocycles and discrete Cantor sets	63
9 The lower bound, completion of the proof	82
10 Further questions and directions	99

1 Introduction

1.1 Historical overview

A simple random walk in a random scenery may be described by the following data:

- The first ingredient is the space $\{\pm 1\}^{\mathbb{Z}}$ with the product measure $\nu_{1/2}^{\otimes \mathbb{Z}}$, where $\nu_{1/2} = \frac{1}{2}(\delta_1 + \delta_{-1})$. This gives the space of possible step-sequences of the walk, from the infinite past and to the infinite future, with the usual i.i.d. law for those steps.
- The second ingredient is a probability-preserving system $(C^{\mathbb{Z}}, \mu, S)$, where C is a finite set of ‘colours’, S is the leftward coordinate-shift on $C^{\mathbb{Z}}$, and μ is an S -invariant probability on $C^{\mathbb{Z}}$. This μ is the law of a random scenery, which decorates every point in \mathbb{Z} with a colour from C .

Elements $(y_n)_n \in \{\pm 1\}^{\mathbb{Z}}$ may be identified bijectively with paths $\mathbb{Z} \rightarrow \mathbb{Z}$ which pass through the origin and whose increments are all -1 or 1 , by identifying y_n with the increment from n to $n + 1$. This converts $(y_n)_n$ into the trajectory taken by the walker, as seen from her current location. The coordinate shift on $(y_n)_n$ acts on this picture by shifting the origin of time, but retaining the feature that the trajectory passes through the origin: that is, we always view the trajectory from the walker’s current location.

This description may naturally be combined with the walker’s view of the scenery to form a probability-preserving system (Z, ρ, R) which captures the whole

of the above picture. First let

$$(Z, \rho) := (\{\pm 1\}^{\mathbb{Z}} \times C^{\mathbb{Z}}, \nu_{1/2}^{\otimes \mathbb{Z}} \otimes \mu).$$

For the dynamics, think of $((y_n)_n, (x_m)_m) \in Z$ as a pair

(trajectory through origin, scenery viewed by walker at origin),

and let R be the transformation which shifts time one step forward, but preserves the feature that the walker's location is the origin. In notation, this is

$$R((y_n)_n, (x_m)_m) = ((y_{n+1})_n, (x_{m+y_0})_m).$$

One checks easily that ρ is R -invariant. This system is called the **random walk in random scenery** μ , and will be denoted RWRS_{μ} .

The systems RWRS_{μ} are important in ergodic theory because they are simple and natural examples of an abstract phenomenon: for many possible choices of μ , they are K-automorphisms but not Bernoulli systems. This was conjectured by Adler, Ornstein and Weiss, who observed that the K-property is fairly easy to prove (it also holds for more complicated random walks, as shown by Meilijson [Mei74]). However, non-Bernoullicity was not proved at that time, and it was recorded as an open problem in [Wei72]. This problem was solved by Kalikow (who refers to this as the ' $[T, T^{-1}]$ system', as have many more recent authors).

Theorem ([Kal82]) *The process $\text{RWRS}_{\nu_{1/2}^{\otimes \mathbb{Z}}}$ is not Bernoulli.* □

The heart of Kalikow's work is to show that $\text{RWRS}_{\nu_{1/2}^{\otimes \mathbb{Z}}}$ does not have the Very Weak Bernoulli property, one of the equivalent characterizations of Bernoullicity involved in Ornstein's famous solution of the Bernoulli Isomorphism Problem.

Other non-Bernoulli K-automorphisms were constructed before Kalikow's work: the first in [Orn73], and then a continuum family of them in [OS73]. However, those examples were all obtained by cutting and stacking for this deliberate purpose.

For any abstract ergodic-theoretic phenomenon, it is of additional interest to find examples that arise naturally from other parts of mathematics (see Section 14 of Thouvenot's essay [Tho02] for further discussion). For non-Bernoulli K-automorphisms, more progress in this direction was made by Feldman in [Fel76]. He exhibited some examples in the form of skew products, somewhat resembling RWRSs, as an application of his new notion of loose Bernoullicity. However, these examples still required the a priori cut-and-stack construction of a non-loosely-Bernoulli automorphism. A smooth version of this construction was then carried

out by Katok in [Kat80b], who points out that Ratner’s formidable work [Rat79] provides natural, geometric non-loosely-Bernoulli transformations for ingredients. Nevertheless, following [Kal82], RWRSs remain the principal ‘natural’ examples of non-Bernoulli K-automorphisms (and, indeed, Kalikow actually shows that $\text{RWRS}_{\nu_{1/2}^{\otimes \mathbb{Z}}}$ is not even loosely Bernoulli, although we will ignore this strengthening here). Thouvenot gives an overview of these developments in [Tho02, Section 12], as well as more complete references.

Since [Kal82], Kalikow’s argument has been generalized in various directions. Smooth examples of skew products analogous to RWRSs are shown to satisfy the same conclusion in [Rud88], following a suggestion in [Kat80b]. In these examples, the trajectories of simple random walk are replaced by the sequences of ergodic sums of a smooth function over an Anosov diffeomorphism. More recently, [dHS97] analysed quite general examples of random walks in \mathbb{Z}^d , $d \geq 1$, with sceneries given by shift-invariant measures on $C^{\mathbb{Z}^d}$, obtaining new non-Bernoulli-K examples when $d = 2$.

(Several more recent works have also explored necessary and sufficient conditions for a RWRS with its obvious generating partition to be Weakly Bernoulli. This is a more restrictive question than Very Weak Bernoullicity, and can be approached using simpler methods than Kalikow’s. However, it is not an invariant of measure-theoretic isomorphism. We will not discuss it further in this paper.)

Having shown that some RWRSs are not Bernoulli, it is natural to ask when they are isomorphic to *each other*. Kalikow’s method does not seem to resolve this question directly, even for i.i.d. sceneries. First, note that if $(C^{\mathbb{Z}}, \mu, S) \cong (D^{\mathbb{Z}}, \theta, S)$, then $\text{RWRS}_{\mu} \cong \text{RWRS}_{\theta}$, since the former isomorphism may simply be applied to the second coordinate of $\{\pm 1\}^{\mathbb{Z}} \times C^{\mathbb{Z}}$. However, the reverse implication can fail. For instance, if $(C^{\mathbb{Z}}, \mu, S)$ is a coding of an ergodic circle-rotation, then RWRS_{μ} is an isometric extension of the Bernoulli shift $(\{\pm 1\}^{\mathbb{Z}}, \nu_{1/2}^{\otimes \mathbb{Z}}, S)$, and in this case it is still Bernoulli of the same entropy [AS72, AS74], so

$$\text{RWRS}_{\text{ergod. rotn.}} \cong \text{RWRS}_{\text{trivial system}}.$$

On the other hand, Kalikow’s result itself shows that the isomorphism class of RWRS_{ν} does remember something (necessarily isomorphism-invariant) about the scenery process.

The purpose of this paper is to show that this includes the entropy of the scenery process. This result seems to have been expected for some time; I learnt of this expectation from J.-P. Thouvenot, but it is also hinted at in Vershik’s paper [Ver00] in connection with his notion of ‘secondary entropy’.

1.2 Statement of the main results

The main result below applies to a generalization of RWRS processes constructed from certain ‘cocycle random walks’. To formulate it, suppose that $\mathbf{Y} = (Y, \nu, S)$ is a probability-preserving system, that $\sigma : Y \rightarrow \mathbb{R}$ is measurable, and that $\mathbf{X} = (X, \mu, T)$ is a jointly measurable and probability-preserving action of \mathbb{R} . Then the **generalized RWRS system with base (\mathbf{Y}, σ) and fibre \mathbf{X}** is the resulting skew-product transformation on $(Y \times X, \nu \otimes \mu)$:

$$(S \times_{\sigma} T)(y, x) := (Sy, T^{\sigma(y)}x).$$

This system will be denoted $\mathbf{Y} \times_{\sigma} \mathbf{X}$.

For example, suppose that

$$\mathbf{Y} := (\{\pm 1\}^{\mathbb{Z}}, \nu_{1/2}^{\otimes \mathbb{Z}}, S), \quad \sigma((y_n)_{n \in \mathbb{Z}}) := y_0,$$

and that (X, μ, T) is a continuous-time flow such that $(X, \mu, T^1) \cong (C^{\mathbb{Z}}, \mu', S)$. Then

$$\mathbf{Y} \times_{\sigma} \mathbf{X} \cong \text{RWRS}_{\mu'}.$$

We shall prove that for certain fixed choices of (\mathbf{Y}, σ) , the entropy of \mathbf{X} is an isomorphism-invariant of the whole generalized RWRS system $\mathbf{Y} \times_{\sigma} \mathbf{X}$. The argument will assume some quite delicate conditions on the system \mathbf{Y} and cocycle σ . In the first place:

$Y \subseteq A^{\mathbb{Z}}$ is a subshift of finite type; S is the coordinate-shift; ν is a Gibbs measure for a Hölder continuous potential on Y ; and σ is a Hölder continuous non-coboundary with $\int \sigma d\nu = 0$.

We refer to these assumptions collectively as (\mathbf{Y}, σ) being a ‘well-distributed pair’. The proofs below will make use of this assumption in many different ways. It could probably be replaced with a longer list of more bespoke assumptions, but it seems simpler to restrict to the above class. Many of the consequences of this assumption that we need assert various kinds of resemblance to Brownian motion at all sufficiently large scales, with some explicit rate on the convergence. This is in a similar spirit to the ‘asymptotically Brownian’ condition required by Rudolph in [Rud88], but technically different.

In addition to the above, we will need to assume that our well-distributed pair satisfies an ‘Enhanced Invariance Principle’, which describes the asymptotic law of the cocycle σ and also its occupation measures over long time-scales. I believe that this principle holds for *all* well-distributed pairs, and can therefore be dropped from explicit mention in Theorem A below. However, it is not yet available in the

literature in that generality. It is available for some more specific examples, and at time of writing I understand that Michael Bromberg is working on the general case. This principle will be formulated carefully in Subsection 3.5.

Theorem A *Suppose that $\mathbf{Y} = (Y, \nu, S)$ and $\sigma : Y \rightarrow \mathbb{R}$ form a well-distributed pair which satisfies the Enhanced Invariance Principle. If \mathbf{X}_i , $i = 1, 2$ are two flows such that there exists a factor map $\mathbf{Y} \ltimes_{\sigma} \mathbf{X}_1 \rightarrow \mathbf{Y} \ltimes_{\sigma} \mathbf{X}_2$, then $h(\mathbf{X}_1) \geq h(\mathbf{X}_2)$.*

Importantly, this allows factor maps that do not act as the identity on the base system \mathbf{Y} . We must therefore find a way to extract the entropy of the scenery from $\mathbf{Y} \ltimes_{\sigma} \mathbf{X}$ as an abstract p.-p. system, without assuming knowledge of the distinguished factor map $\mathbf{Y} \ltimes_{\sigma} \mathbf{X} \rightarrow \mathbf{Y}$.

It is important that one fix the choice of (\mathbf{Y}, σ) . Indeed, the invariant that we shall actually produce takes the form $f(\sigma)h(\mathbf{X})$, where f is some function of σ which is homogeneous of order 1. It is easy to see that if one replaces σ with 2σ and \mathbf{X} with its slowdown by a factor of 2, then the resulting generalized RWRS systems are isomorphic, so this fixing of σ is essential.

Aaronson's recent work [Aar12] implies a special case of Theorem A in which the factor map is assumed to respect the coordinate factor map to \mathbf{Y} . Applied to our setting, Corollary 5 of that paper shows that if \mathbf{Y} and σ are the process and cocycle of classical simple random walk, then a relative factor map

$$\begin{array}{ccc} \mathbf{Y} \ltimes_{\sigma} \mathbf{X}_1 & \xrightarrow{\quad} & \mathbf{Y} \ltimes_{\sigma} \mathbf{X}_2 \\ & \searrow \text{coord. proj.} & \swarrow \text{coord. proj.} \\ & \mathbf{Y} & \end{array}$$

can exist only if $h(\mathbf{X}_1) \geq h(\mathbf{X}_2)$. (Aaronson also handles the case of other stable random walks, which we leave aside here.) For a canonical choice of generating partition \mathcal{R} for these systems $\mathbf{X} \ltimes_{\sigma} \mathbf{Y}$, this result follows from a calculation of the distributions of the relative complexities of (\mathcal{R}, N) -names over the base system \mathbf{Y} , regarded as random variable on the probability space (Y, ν) . Our work below will turn out to need many of the same calculations as Aaronson's. However, relative complexities give an invariant only of relative isomorphism: they do not serve to control arbitrary factor maps $\mathbf{Y} \ltimes_{\sigma} \mathbf{X}_1 \rightarrow \mathbf{Y} \ltimes_{\sigma} \mathbf{X}_2$.

Theorem A also has precedents in the study of non-invertible RWRS processes, for which a point in the state space records only the future trajectory of the walk. The analog of Theorem A with one-sided simple random walk in the base was proved by Heicklen, Hoffman and Rudolph in [HHR00], and a generalization to some other skew products, including some smooth examples, was given by Ball in [Bal03]. In some ways the steps in our work below reflect those papers, except

that they make essential use of some extra isomorphism-invariant structure of a non-invertible transformation: the decreasing filtration of pre-images of the σ -algebra. This idea goes back to work of Vershik around 1970: see [Ver94].

Conditionally on the Enhanced Invariance Principle, Theorem A also covers certain smooth analogs of RWRSs on compact manifolds, using appropriate codings from Gibbs measures on subshifts. Arguably, these form the most ‘natural’ among all kinds of example in ergodic theory (see again [Tho02, Section 14], which includes a discussion of some of these smooth RWRS-like examples, as studied in [Rud88]). The first smooth non-Bernoulli K-automorphisms were constructed in [Kat80b], but again with a more complicated description.

For instance, let $A : \mathbb{T}^2 \longrightarrow \mathbb{T}^2$ be a hyperbolic toral automorphism and m be the Haar probability measure, let $g = (g^t)_{t \in \mathbb{R}}$ be an Anosov flow on a compact manifold M that preserves the Riemannian volume-form μ , and on the space $(\mathbb{T}^2 \times M, m \otimes \mu)$ consider the skew-product transformations

$$T_r(x_1, x_2, p) := (A(x_1, x_2), g^{r \sin x_1} p).$$

Kalikow’s argument itself was extended to cover examples such as these in [Rud88].

These can clearly be written as skew products of the form $\mathbf{Y} \times_{\sigma} \mathbf{X}$ with $\mathbf{Y} = (\mathbb{T}^2, m, A)$ and $\mathbf{X} = (M, \mu, g)$. This \mathbf{Y} has a coding given by an a.e. one-one Hölder function $F : (Y, \nu, S) \longrightarrow (\mathbb{T}^2, m, A)$ for some SFT Y and Hölder-potential Gibbs measure ν (see [Bow08]). Therefore Theorem A applies to these examples provided one knows the Enhanced Invariance Principle. Since $A \times_{r\sigma} g = A \times_{\sigma} g^{(r)}$, where $g^{(r)}$ is the speedup of g by the constant factor r , it follows that the quantity

$$h(\mu, g^{(r)}) = r h(\mu, g)$$

is an isomorphism-invariant of $\mathbf{Y} \times_{\sigma} \mathbf{X}$. Since $h(\mu, g)$ is finite and positive, these values are distinct for distinct r , and so, conditionally on the Enhanced Invariance Principle, this family of examples proves the following.

Conditional Corollary B *For any $h \in (0, \infty)$, there is a compact manifold with a smooth volume form that admits continuum-many smooth, volume-preserving K-automorphisms of entropy h which are pairwise non-isomorphic.* \square

The corresponding result for non-invertible maps was also proved by Ball in [Bal03]. This possible consequence of the current work was brought to my attention by J.-P. Thouvenot.

1.3 A new isomorphism invariant

The key to Theorem A will be a new isomorphism-invariant of probability-preserving systems.

The definition of this new invariant is rather involved, and will not be given in full until Section 5. However, some motivation for it can be given in advance. This will involve standard notions from information theory, which the unfamiliar reader can find recalled in Subsection 2.3.

Consider again the basic examples RWRS_μ . First, let us recall why the Kolmogorov-Sinai entropy of RWRS_μ does not give any information about $h(\mu, S)$. Let $\mathbf{Y} := (\{\pm 1\}^\mathbb{Z}, \nu_{1/2}^{\otimes \mathbb{Z}}, S)$, let $\mathbf{X} := (C^\mathbb{Z}, \mu, S)$ be a scenery process (here in discrete time), let $\rho = \nu_{1/2}^{\otimes \mathbb{Z}} \otimes \mu$, and let $\sigma : \{\pm 1\}^\mathbb{Z} \longrightarrow \{\pm 1\}$ be the time-zero coordinate. These data together define $\text{RWRS}_\mu = \mathbf{Y} \ltimes_\sigma \mathbf{X}$. Let $(Z, \rho) := (\{\pm 1\}^\mathbb{Z} \times C^\mathbb{Z}, \nu_{1/2}^{\otimes \mathbb{Z}} \otimes \mu)$, and let

$$\alpha : Z \longrightarrow \{\pm 1\} \times C$$

be the time-zero map corresponding to the obvious generating partition \mathcal{R} for RWRS_μ . Let \mathcal{Q} be the time-zero partition of $C^\mathbb{Z}$.

For $N \in \mathbb{N}$, let $\rho_N := \alpha_*^{[0;N]} \mu \in \text{Pr}((\{\pm 1\} \times C)^N)$ be the distribution of the (α, N) -name

$$\begin{aligned} & (\alpha(z), \alpha((S \ltimes_\sigma S)(z)), \dots, \alpha((S \ltimes_\sigma S)^{N-1}(z))) \\ &= ((y_0, \dots, y_{N-1}), (x_0, x_{y_0}, x_{\sigma_1^y}, \dots, x_{\sigma_{N-1}^y})) \end{aligned}$$

when $z = (y, x)$ is drawn from ρ . The Kolmogorov-Sinai entropy of RWRS_μ is given by the leading-order behaviour of the sequence of Shannon entropies $H(\rho_N) = H_\rho(\mathcal{R}^{[0;N]})$.

This can be computed in terms of the information function of $\mathcal{R}^{[0;N]}$:

$$I_{\mu, \mathcal{R}^{[0;N]}} : Z \longrightarrow [0, \infty) : z \mapsto -\log \rho(\mathcal{R}^{[0;N]}(z)).$$

In the N -name written above, the string (y_0, \dots, y_{N-1}) is equally likely to be any element of $\{\pm 1\}^N$, so this contributes $(\log 2)N$ to the value $I_{\rho, \mathcal{R}^{[0;N]}}(y, x)$. However, having fixed (y_0, \dots, y_{N-1}) , the possible output strings $(x_0, x_{y_0}, \dots, x_{\sigma_{N-1}^y})$ are in bijective correspondence with the scenery-portions $(x_m)_{m \in \sigma_{[0;N]}^y}$, where $\sigma_{[0;N]}^y = \{\sigma_n^y \mid n \in [0; N]\}$. This gives the total value for the information function as

$$I_{\rho, \mathcal{R}^{[0;N]}}(y, x) = (\log 2)N + I_{\mu, \mathcal{Q}^{\sigma_{[0;N]}^y}}(x).$$

By the Shannon-McMillan Theorem (recalled as Theorem 3.1 below), for typical (y, x) and large N this is

$$(\log 2)N + h(\mu, S)|\sigma_{[0;N]}^y| + o(|\sigma_{[0;N]}^y|). \quad (1)$$

Simple random walk on \mathbb{Z} behaves diffusively, meaning that for typical y the cardinality $|\sigma_{[0;N]}^y|$ is of order \sqrt{N} . Therefore for typical y the above value is given by

$$(\log 2)N + h(\mu, S)c_N(y)\sqrt{N} + o(\sqrt{N})$$

for some value $c_N(y)$ which is typically of order 1.

Thus, if we ignore certain rare events in (y, x) , then the entropy of the scenery contributes only a correction of order \sqrt{N} to

$$H_\rho(\mathcal{R}^{[0;N]}) = \int I_{\rho, \mathcal{R}^{[0;N]}}(z) \rho(dz).$$

This sublinear correction disappears in the limit that computes $h(\text{RWRS}_\mu)$.

In general, sublinear terms in the growth-rate of $H_\rho(\mathcal{R}^{[0;N]})$ are not isomorphism-invariant, so we cannot use the above calculation to prove the invariance of $h(\mu, S)$. Towards fixing this problem, let us next consider a different way to look at these corrections, in terms of another information-theoretic quantity: the mutual information between the N -step past $\mathcal{R}^{[-N;0]}$ and the N -step future $\mathcal{R}^{[0;N]}$. By definition, this is

$$H_\rho(\mathcal{R}^{[-N;0]}) + H_\rho(\mathcal{R}^{[0;N]}) - H_\rho(\mathcal{R}^{[-N;N]}).$$

Now each term here may be written as an integral of information functions:

$$\int (I_{\rho, \mathcal{R}^{[-N;0]}}(z) + I_{\rho, \mathcal{R}^{[0;N]}}(z) - I_{\rho, \mathcal{R}^{[-N;N]}}(z)) \rho(dz).$$

Let us again ask about the typical behaviour of the integrand here for $z \sim \rho$, ignoring certain extreme events (specifically, that the simple random walk covers much more ground than expected between times $-N$ and N).

Substituting from (1), we find that for typical $z = (y, x)$ and sufficiently large N we have

$$\begin{aligned} & I_{\rho, \mathcal{R}^{[-N;0]}}(z) + I_{\rho, \mathcal{R}^{[0;N]}}(z) - I_{\rho, \mathcal{R}^{[-N;N]}}(z) \\ &= (\log 2)(N + N - 2N) + h(\mu, S)(|\sigma_{[-N;0]}^y| + |\sigma_{[0;N]}^y| - |\sigma_{[-N;N]}^y|) + o(\sqrt{N}) \\ &= 0 + h(\mu, S)|\sigma_{[-N;0]}^y \cap \sigma_{[0;N]}^y| + o(\sqrt{N}). \end{aligned}$$

Heuristically, this calculation runs as follows: the steps taken by the walk in the past and future are independent, so contribute nothing to the mutual information;

and the remaining mutual information is all contributed by *that portion of the scenery visited by both the N -step past and the N -step future*.

Now, an easy appeal to Donsker's Invariance Principle gives that as $N \rightarrow \infty$, the random variable $y \mapsto |\sigma_{[-N;0]}^y \cap \sigma_{[0;N]}^y|/\sqrt{N}$ converges in law to the random variable $\mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]})$, where B and B' are independent Brownian motions. This suggests that, provided one allows for this limiting behaviour of the random variable $y \mapsto |\sigma_{[-N;0]}^y \cap \sigma_{[0;N]}^y|$, the constant $h(\mu, S)$ should be visible in the asymptotic behaviour of $I_\rho(\mathcal{R}^{[-N;0]}; \mathcal{R}^{[0;N]})$ (perhaps after allowing the excision of a small-measure subset of $\{\pm 1\}^{\mathbb{Z}} \times C^{\mathbb{Z}}$ to remove 'pathological' random-walk trajectories).

As with the sublinear entropy-corrections themselves, one expects that the sequence of mutual informations $I_\rho(\mathcal{R}^{[-N;0]}; \mathcal{R}^{[0;N]})$ does not give an isomorphism-invariant of general processes $(\mathbf{Z}, \mathcal{R})$ (although I have not proved this carefully). The key remaining idea is to modify the definition of I_ρ to obtain a more robust quantity.

The way to do this is suggested by a general viewpoint that already has already been very fruitful in the study of Kolmogorov-Sinai entropy. Given an ergodic system $\mathbf{Z} = (Z, \rho, R)$, a finite measurable partition \mathcal{R} of Z and a finite-valued map $\alpha : Z \rightarrow A$ which generates \mathcal{R} , the Shannon-McMillan Theorem expresses $h(\mathbf{Z}, \mathcal{R})$ as the exponential growth rate of the effective number of points in A^N needed to support $\alpha_*^{[0;N]} \rho$. However, as observed by Feldman ([Fel80]), this may also be approximated by choosing some sufficiently small $\delta > 0$, and then asking after the exponential growth rate of the number of (δN) -balls needed to cover most of the measure $\alpha_*^{[0;N]} \rho$ in the Hamming metric spaces

$$(A^N, d_{\text{Ham}}).$$

Having proved this covering-number representation, the isomorphism-invariance of $h(\mathbf{Z}, \mathcal{R})$ follows fairly easily, since an isomorphism of processes may be approximated, for sufficiently large N , by a sequence of *Lipschitz* maps between these metric spaces, for which the change in those covering numbers is easily controlled.

Inspired by this viewpoint, our replacement for the sequence $I_\rho(\mathcal{R}^{[-N;0]}; \mathcal{R}^{[0;N]})$ will be a sequence of values measuring how much 'information' is held by both of the partitions $\mathcal{R}^{[-N;0]}$ and $\mathcal{R}^{[0;N]}$ if one insists that this 'information' can be recovered robustly if one allows small errors according to the Hamming metrics on $A^{[-N;0]}$ and $A^{[0;N]}$.

An important step in this paper is the rigorous development of this new invariant, via notions defined on abstract spaces that carry pairs of metrics. This will be the work of Section 5.

1.4 Outline of the remaining sections

Sections 2 and 3 present a variety of standard or routine results that will be needed later, concerning analysis and dynamics respectively. Subsection 3.5 formulates the Enhanced Invariance Principle and describes some cases in which it is known.

Section 4 is a warm-up for the rest of the paper. It describes some basic features of the marginal metric spaces that arise from the skew-products in Theorem A.

Section 5 introduces the specific new isomorphism-invariant at the heart of the proof of Theorem A, estimates it in a few simple cases, and states the more precise Theorem 5.23 about its behaviour for the skew-products that appear in Theorem A.

Sections 6 and 7 prove the upper bound asserted in Theorem 5.23.

Section 8 returns to the study of well-distributed cocycles, focusing on some more subtle properties that are needed for the lower bound. Chief among these is the ability, for a ‘typical’ trajectory of the cocycle σ over the interval $\{0, 1, \dots, N-1\}$, to find very many somewhat large subsets of this interval on which σ is injective, and which have a discrete ‘Cantor-like’ structure.

Section 9 then uses these finer properties to prove the lower bound asserted in Theorem 5.23, and hence complete the proof of Theorem A. This is more difficult than the upper-bound proof, and draws important ideas from [Kal82].

Finally, Section 10 formulates some open questions and directions for further investigation.

Acknowledgements

Jean-Paul Thouvenot shared with me several important insights in connection with the problems addressed here, as well as his enthusiasm for them. Peter Nandori introduced me to Local Limit Theorems for smooth cocycles and helped me greatly in understanding them, and then Michael Bromberg shared with me his current progress on local times for cocycles over Gibbs-Markov processes. \triangleleft

2 Preliminaries: analysis and probability

2.1 Basic conventions

An **interval** will be either an interval in \mathbb{R} or a discrete interval in \mathbb{Z} ; the ambient set will always be clear from the context. If $a, b \in \mathbb{Z}$ with $a \leq b$ then $[a; b] := [a; b+1) := (a-1; b] := \{a, a+1, \dots, b\}$. Sometimes we use the abbreviation $[n] := [0; n)$.

Given an interval $K \subseteq \mathbb{R}$, we will let $\text{Int}(K)$ denote the collection of all nonempty compact subintervals of K . We give it the topology inherited from the obvious identification with $\{(u, v) \in K \mid u \leq v\} \subseteq \mathbb{R}^2$.

Lebesgue measure on \mathbb{R} will be denoted by \mathcal{L}^1 . If I is a bounded interval in either \mathbb{R} or \mathbb{Z} , then U_I will denote the uniform probability distribution on I .

We will use \star to denote convolution of functions or measures on \mathbb{R} , in any case in which it is well-defined.

In this paper, a **mollifier** will be a compactly-supported smooth function $\varphi : \mathbb{R} \rightarrow [0, \infty)$ which is symmetric about the origin and satisfies $\int \varphi d\mathcal{L}^1 = 1$.

The following popular notation from harmonic analysis will be useful later. Given two collections $(A_i)_{i \in I}, (B_i)_{i \in I}$ of non-negative real numbers and another structure or quantity X , we write $A_i \lesssim_X B_i$ to assert that there is a constant $C \in (0, \infty)$ depending only on X such that $A_i \leq CB_i$ for all i . We write $A_i \sim_X B_i$ in case both $A_i \lesssim_X B_i$ and $B_i \lesssim_X A_i$.

2.2 Probability

Various later arguments will involve comparisons with Brownian motion. We will always let $W \in \text{Pr } C[0, \infty)$ be the classical Wiener measure, and let $W_{[0,1]} \in \text{Pr } C[0, 1]$ be the law of $B|_{[0,1]}$ for $B \sim W$. This latter is supported on the closed subset

$$C_0(0, 1] := \{f \in C[0, 1] \mid f(0) = 0\}.$$

If (X, Σ, μ) is a probability space and $A \in \Sigma$ has $\mu(A) > 0$, then $\mu|_A$ will denote the conditional measure $\mu(A \cap \cdot)/\mu(A)$.

In our dynamical applications, all probability spaces will be standard Borel, and we will generally omit their σ -algebras from the notation.

We will later make several uses of the following quantitative approximation to absolute continuity.

Definition 2.1 (Approximate absolute continuity). *Let (X, Σ) be a measurable space, μ and ν be finite measures on X , and $\varepsilon \in [0, \infty)$ and $M \in (0, \infty)$. Then we write that $\mu \ll_{M, \varepsilon} \nu$ if*

$$\mu(A) \leq M\nu(A) + \varepsilon \quad \forall A \in \Sigma,$$

and we write that $\mu \sim_{M, \varepsilon} \nu$ if $\mu \ll_{M, \varepsilon} \nu$ and $\nu \ll_{M, \varepsilon} \mu$.

In case μ and ν are both probability measures, an easy exercise gives

$$\mu \ll_{1, \varepsilon} \nu \iff \nu \ll_{1, \varepsilon} \mu \iff \mu \sim_{1, \varepsilon} \nu$$

(where the first equivalence holds because the Jordan decomposition gives $(\mu - \nu)^+(X) = (\mu - \nu)^-(X)$ for any two probability measures). On the other hand, $\mu \ll_{M, 0} \nu$ if and only if μ is absolutely continuous with respect to ν and $\|\mathrm{d}\mu/\mathrm{d}\nu\|_{L^\infty(\nu)} \leq M$.

The following basic properties are also routine to verify.

Lemma 2.2. *Approximate absolute continuity enjoys the following properties:*

- If $\mu_1 \ll_{M_1, \varepsilon_1} \mu_2$ and $\mu_2 \ll_{M_2, \varepsilon_2} \mu_3$, then

$$\mu_1 \ll_{M_1 M_2, M_1 \varepsilon_2 + \varepsilon_1} \mu_3.$$

- If $\mu, \nu, \theta \in \text{Pr } \mathbb{R}$ and $\mu \ll_{M, \varepsilon} \nu$, then also $\theta \star \mu \ll_{M, \varepsilon} \theta \star \nu$. □

2.3 Information Theory

We shall make use of several notions from Information Theory. The main definitions are recalled here, but we shall largely take standard facts for granted: Cover and Thomas [CT06] is a canonical reference.

Given a countable set A and $\mu \in \text{Pr } A$, the **Shannon entropy** of μ is

$$H(\mu) := - \sum_{a \in A} \mu\{a\} \log \mu\{a\} \in [0, +\infty].$$

Relatedly, if (X, μ) is any probability space and $\varphi : X \rightarrow A$ is measurable, then $H_\mu(\varphi) := H(\varphi_* \mu)$; and if \mathcal{P} is a countable measurable partition of X , then $H_\mu(\mathcal{P}) := H_\mu(\varphi)$ for any choice of countable-valued map φ whose level-sets are the cells of \mathcal{P} .

If (X, Σ, μ) is any probability space and ν is another probability on X , then the **Kullback-Leibler divergence of ν with respect to μ** is

$$D_{\text{KL}}(\nu | \mu) := \begin{cases} \int_X \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu & \text{if } \nu \ll \mu \\ +\infty & \text{else} \end{cases} \in [0, +\infty].$$

Next, suppose that \mathcal{P} and \mathcal{Q} are two countable measurable partitions of (X, Σ, μ) . Then the **conditional entropy** of \mathcal{P} given \mathcal{Q} is the quantity

$$H_\mu(\mathcal{P} | \mathcal{Q}) := \sum_{C \in \mathcal{Q}} \mu(C) H_{\mu|_C}(\mathcal{P})$$

(where we interpret those $C \in \mathcal{Q}$ for which $\mu(C) = 0$ as contributing zero). The **mutual information** of \mathcal{P} and \mathcal{Q} under μ is defined by

$$I_\mu(\mathcal{P}; \mathcal{Q}) := H_\mu(\mathcal{P}) - H_\mu(\mathcal{P} | \mathcal{Q}).$$

A standard calculation shows that this is symmetric in \mathcal{P} and \mathcal{Q} , and also that

$$\begin{aligned} I_\mu(\mathcal{P}; \mathcal{Q}) &= H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}) - H_\mu(\mathcal{P} \vee \mathcal{Q}) \\ &= \int D_{\text{KL}}(\varphi_*(\mu|_{\mathcal{Q}(x)}) | \varphi_* \mu) \mu(dx), \end{aligned} \tag{2}$$

where $\varphi : X \longrightarrow A$ is any finite-valued function generating the partition \mathcal{P} (see, for instance, Equations (2.45) and (2.36) in [CT06, Section 2.4]). More generally, given a third partition \mathcal{R} , the **conditional mutual information** of \mathcal{P} and \mathcal{Q} given \mathcal{R} is

$$\begin{aligned} I_\mu(\mathcal{P}; \mathcal{Q} | \mathcal{R}) &:= H_\mu(\mathcal{P} | \mathcal{R}) - H_\mu(\mathcal{P} | \mathcal{Q} \vee \mathcal{R}) \\ &= \int I_{\mu|_{\mathcal{R}(x)}}(\mathcal{P}; \mathcal{Q}) \mu(dx), \end{aligned}$$

where the second equality is another standard calculation.

These definitions easily give the following.

Lemma 2.3. *Let (X_i, Σ_i, μ_i) for $i = 1, 2$ be probability spaces, and for each i let $\mathcal{P}_i, \mathcal{Q}_i$ and \mathcal{R}_i be countable measurable partitions of X_i . Then*

$$I_{\mu_1 \otimes \mu_2}(\mathcal{P}_1 \otimes \mathcal{P}_2; \mathcal{Q}_1 \otimes \mathcal{Q}_2 | \mathcal{R}_1 \otimes \mathcal{R}_2) = I_{\mu_1}(\mathcal{P}_1; \mathcal{Q}_1 | \mathcal{R}_1) + I_{\mu_2}(\mathcal{P}_2; \mathcal{Q}_2 | \mathcal{R}_2).$$

□

We will also need the following simple but less standard calculations.

Lemma 2.4 (Conditioning mutual information on a subset). *If (X, Σ, μ) is a probability space, \mathcal{P}, \mathcal{Q} , and \mathcal{R} are countable partitions in Σ , and $A \in \Sigma$ has positive measure, then*

$$\mu(A) I_{\mu|_A}(\mathcal{P}; \mathcal{Q} | \mathcal{R}) \leq \log 2 + I_\mu(\mathcal{P}; \mathcal{Q} | \mathcal{R}).$$

Proof. Let $\mathcal{A} := \{A, X \setminus A\}$. From the definition of conditional mutual information and the fact that it is always non-negative ([CT06, Corollary 2.6.3]), one obtains

$$\begin{aligned} \mu(A) I_{\mu|_A}(\mathcal{P}; \mathcal{Q} | \mathcal{R}) &\leq \mu(A) I_{\mu|_A}(\mathcal{P}; \mathcal{Q} | \mathcal{R}) + \mu(X \setminus A) I_{\mu|_{X \setminus A}}(\mathcal{P}; \mathcal{Q} | \mathcal{R}) \\ &= I_\mu(\mathcal{P}; \mathcal{Q} | \mathcal{R} \vee \mathcal{A}). \end{aligned}$$

The Chain Rule for mutual information ([CT06, Theorem 2.5.2]) gives

$$\begin{aligned} I_\mu(\mathcal{P} \vee \mathcal{A}; \mathcal{Q} | \mathcal{R}) &= I_\mu(\mathcal{A}; \mathcal{Q} | \mathcal{R}) + I_\mu(\mathcal{P}; \mathcal{Q} | \mathcal{R} \vee \mathcal{A}) \\ &\implies I_\mu(\mathcal{P}; \mathcal{Q} | \mathcal{R} \vee \mathcal{A}) \leq I_\mu(\mathcal{P} \vee \mathcal{A}; \mathcal{Q} | \mathcal{R}), \end{aligned}$$

and now another use of the definitions, subadditivity of entropy and the Data-Processing Inequality gives

$$\begin{aligned} I_\mu(\mathcal{P} \vee \mathcal{A}; \mathcal{Q} | \mathcal{R}) &= H_\mu(\mathcal{P} \vee \mathcal{A} | \mathcal{R}) - H_\mu(\mathcal{P} \vee \mathcal{A} | \mathcal{Q} \vee \mathcal{R}) \\ &\leq H_\mu(\mathcal{A}) + H_\mu(\mathcal{P} | \mathcal{R}) - H_\mu(\mathcal{P} | \mathcal{Q} \vee \mathcal{R}) \\ &= H_\mu(\mathcal{A}) + I_\mu(\mathcal{P}; \mathcal{Q} | \mathcal{R}). \end{aligned}$$

Finally, $H_\mu(\mathcal{A}) \leq \log |\mathcal{A}| = \log 2$. \square

Lemma 2.5 (Uniform integrability from relative entropy bound). *If (X, Σ, μ) is a probability space and $\nu = f \cdot \mu \in \text{Pr } X$ with $D := D_{\text{KL}}(\nu | \mu) < \infty$, then for any $C > 0$ one has*

$$\nu \ll_{e^C, (D+e^{-1})/C} \mu.$$

Proof. Since $D = \int f \log f \, d\mu$, and the function $t \mapsto t \log t$ has a global minimum at $t = e^{-1}$ with value $-e^{-1}$, one has

$$\begin{aligned} \nu\{f > e^C\} &= \int_{\{f > e^C\}} f \, d\mu = \int_{\{\log f > C\}} f \, d\mu \leq \int_{\{\log f > C\}} f \frac{\log f}{C} \, d\mu \\ &\leq \frac{1}{C} \int |f \log f| \, d\mu \leq \frac{D + e^{-1}}{C}. \end{aligned}$$

Therefore for any measurable $A \subseteq X$ one has

$$\nu(A) = \int_A f \, d\mu \leq e^C \mu(A) + \nu(A \cap \{f > e^C\}) \leq e^C \mu(A) + \frac{D + e^{-1}}{C}.$$

\square

2.4 Metric and pseudometric spaces

If (X, d) is a metric or pseudometric space, $x \in X$ and $r \geq 0$, then

$$B_r^d(x) := \{y \in X \mid d(x, y) < r\}$$

is the radius- r open ball around x . It will sometimes be abbreviated to $B_r(x)$ if d is understood. If $F \subseteq X$, then $B_r(F) := \bigcup_{x \in F} B_r(x)$. A subset $F \subseteq X$ is **r -separated** if

$$d(x, y) \geq r \quad \forall x, y \in F \text{ distinct.}$$

The **r -covering number** of (X, d) is

$$\text{cov}((X, d), r) = \min\{|F| \mid F \subseteq X, B_r(F) = X\}.$$

A **metric measure** (**‘m.m.’**) **space** is a triple (X, d, μ) consisting of a metric space (X, d) and a Radon measure μ on X . In this paper it will always be tacitly assumed that μ is finite. If $\mu(X) = 1$ then (X, d, μ) is a **metric probability** (**‘m.p.’**) **space**. All m.m. spaces appearing below will either be compact or arise as Borel subsets of compact spaces.

It will sometimes be necessary to generalize this class to include pseudometrics. However, the open balls for a pseudometric may not generate the whole of the relevant σ -algebra. Thus, in this paper, a **pseudometric measure** (resp. **pseudometric probability**) (**'psm.m.'**, resp. **'psm.p'**) **space** will be a triple (X, d, μ) in which X is a standard Borel space, μ is a measure (resp. probability) on X , and $d : X \times X \longrightarrow [0, \infty)$ is a pseudometric which is Borel measurable on $X \times X$ and is totally bounded. Clearly all compact m.m. spaces fall into this class. This definition is similar to, though slightly more restrictive than, Vershik's class of 'admissible' pseudometrics in [Ver10].

If X is a standard Borel space, then one may obtain a totally bounded Borel pseudometric d on X by letting (Z, d^Z) be a compact metric space and $\varphi : X \longrightarrow Z$ a Borel map, and then taking $d := d^Z \circ \varphi^{\times 2}$. An easy exercise shows that every totally bounded Borel pseudometric d on X arises this way, by letting (Z, d^Z) be the completion of the quotient of X by the zero-distance equivalence relation defined by d .

If (X, d, μ) is a psm.m. space with σ -algebra Σ and $U \in \Sigma$, then we usually abbreviate

$$(U, d|_{U \times U}, \mu|_{\Sigma \cap U}) =: (U, d, \mu),$$

so this latter has total mass $\mu(U)$. On the other hand, if (X, d, μ) is a psm.p. space and $U \in \Sigma$ has $\mu(U) > 0$, then

$$(U, d|_{U \times U}, \mu(U)^{-1} \cdot \mu|_{\Sigma \cap U}) =: (U, d, \mu|_U),$$

another psm.p. space.

Given a psm.m. space (X, d, μ) and $a, r > 0$, the **a -partial r -covering number** is

$$\text{cov}_a((X, d, \mu), r) := \min\{|F| \mid F \subseteq X, \mu(B_r(F)) > a\}. \quad (3)$$

Much of the work later will concern a natural 'roughening' of the class of Lipschitz maps. Given pseudometric spaces (X, d^X) and (Y, d^Y) , and also $c, L \geq 0$, a map $f : X \longrightarrow Y$ is **c -almost L -Lipschitz** if

$$d^Y(f(x), f(x')) \leq L d^X(x, x') + c \quad \forall x, x' \in X.$$

This class of maps already has a natural place in the study of concentration of measure. For instance, it appears repeatedly in Chapter 3 $\frac{1}{2}$ of Gromov [Gro01] (starting in the proof of 3 $\frac{1}{2}$.15(b)), under the terminology ' K -Lipschitz up to c '.

3 Preliminaries: ergodic theory

We shall need to call on a variety of classical results from ergodic theory, and especially from entropy and Ornstein theory for probability-preserving transformations. Two standard references that emphasize the material we need are Shields [Shi96] and Kalikow and McCutcheon [KM10].

3.1 Probability-preserving systems and their entropy

In the following, a **probability-preserving** ('p.-p.') **system** is a triple (X, μ, T) in which (X, μ) is a standard Borel probability space and $T : X \rightarrow X$ is measurable, has a measurable inverse, and preserves μ . Similarly, a **p.-p. flow** is a triple (X, μ, T) in which (X, μ) is standard Borel and $T : \mathbb{R} \curvearrowright X$ is jointly measurable and μ -preserving. Many properties of such a flow are closely related to properties of its **time-1 system** (X, μ, T^1) .

The classical entropy theory of p.-p. systems is most easily introduced in terms of finite partitions of X (or, equivalently, finite-valued measurable functions on X). We will assume this theory as it is presented, for example, in [Shi96] or [KM10].

An essential tool will be the Shannon-McMillan Theorem. Some further notation will be useful. Suppose that (X, μ, T) is a p.-p. system and that \mathcal{P} is a finite Borel partition of X . A pair such as $(\mathbf{X}, \mathcal{P})$ will be called a **process**. For any subset $F \subseteq \mathbb{Z}$, let

$$\mathcal{P}^F := \bigvee_{n \in F} T^{-n}(\mathcal{P}),$$

where this is interpreted as a new partition in case F is finite, or, more generally, as a σ -subalgebra of the σ -algebra of X if F is infinite. Now let

$$X_{I,\varepsilon}^{\text{SM}} := \{x \in X \mid e^{-(h(\mathbf{X}, \mathcal{P}) + \varepsilon)|I|} < \mu(\mathcal{P}^I(x)) < e^{-(h(\mathbf{X}, \mathcal{P}) - \varepsilon)|I|}\}$$

(so this depends on \mathcal{P} , although the notation suppresses that dependence). Clearly $X_{I+n,\varepsilon}^{\text{SM}} = T^n(X_{I,\varepsilon}^{\text{SM}})$ for every $n \in \mathbb{Z}$.

The following can be found in [KM10, Section 4.2] or [Shi96, Sections I.5 and I.6].

Theorem 3.1 (Shannon-McMillan Theorem). *If (X, μ, T) is ergodic then*

$$\mu(X_{I,\varepsilon}^{\text{SM}}) \rightarrow 1$$

as $|I| \rightarrow \infty$ for any fixed $\varepsilon > 0$. □

The following is also essentially a standard result.

Lemma 3.2. *For any $N \geq 1$, the Kolmogorov-Sinai entropy satisfies*

$$H_\mu(\mathcal{P}^{[0;N]} \mid \mathcal{P}^{[-M;0]}) \downarrow h(\mathbf{X}, \mathcal{P})N \quad \text{as } M \rightarrow \infty.$$

Proof. When $N = 1$, this can be obtained from [KM10, Corollary 423 and Theorem 434] or from [Shi96, Equation I.6(3)]. For general N , the chain rule for relative entropy (see, for instance, [Shi96, Equation I.6(1)] or [CT06, Section 2.5]) gives

$$H_\mu(\mathcal{P}^{[0;N]} \mid \mathcal{P}^{[-M;0]}) = \sum_{n=0}^{N-1} H_\mu(T^{-n}(\mathcal{P}) \mid \mathcal{P}^{[-M;n]}).$$

Since N is fixed, we may now apply the special case to each right-hand summand separately as $M \rightarrow \infty$. \square

Now consider two discrete intervals $I, J \subseteq \mathbb{Z}$ such that $I \cup J$ is also a discrete interval: thus, either one of them is empty, or they are adjacent, or they intersect.

Lemma 3.3. *Given \mathbf{X} and \mathcal{P} , there is a function $g : \mathbb{N} \rightarrow [0, \infty)$ with $g(m) = o(m)$ as $m \rightarrow \infty$ such that*

$$I_\mu(\mathcal{P}^J; \mathcal{P}^I \mid \mathcal{P}^{I \cap J}) \leq g(|J \setminus I|).$$

Proof. The definition of I_μ gives

$$\begin{aligned} I_\mu(\mathcal{P}^J; \mathcal{P}^I \mid \mathcal{P}^{I \cap J}) &= H_\mu(\mathcal{P}^J \mid \mathcal{P}^{I \cap J}) - H_\mu(\mathcal{P}^J \mid \mathcal{P}^I) \\ &= H_\mu(\mathcal{P}^{J \setminus I} \mid \mathcal{P}^{I \cap J}) - H_\mu(\mathcal{P}^{J \setminus I} \mid \mathcal{P}^I), \end{aligned} \quad (4)$$

since $\mathcal{P}^I \vee \mathcal{P}^{I \cap J} = \mathcal{P}^I$. Various cases are now trivial: if either I or J is empty, or if either $I \subseteq J$ or $J \subseteq I$, then this right-hand side collapses to zero.

In the remaining case, we observe that $J \setminus I$ is also a nonempty interval. In this case, standard monotonicity properties of conditional entropy together with Lemma 3.2 give

$$h(\mathbf{X}, \mathcal{P})|J \setminus I| \leq H_\mu(\mathcal{P}^{J \setminus I} \mid \mathcal{P}^I) \leq H_\mu(\mathcal{P}^{J \setminus I} \mid \mathcal{P}^{I \cap J}) \leq H_\mu(\mathcal{P}^{J \setminus I}).$$

However, the right-hand quantity here is of the form

$$h(\mathbf{X}, \mathcal{P})|J \setminus I| + g(|J \setminus I|)$$

for some sublinear function g , so the right-hand side of (4) is bounded by this g , completing the proof. \square

Given a p.-p. transformation $\mathbf{X} = (X, \mu, T)$ and a finite Borel partition \mathcal{P} , one may always choose a finite set A and function $\varphi_0 : X \rightarrow A$ which generates \mathcal{P} . Having done so, let $\varphi_n := \varphi_0 \circ T^n$ for each $n \in \mathbb{Z}$, and more generally $\varphi_F := (\varphi_n)_{n \in F} : X \rightarrow A^F$ for $F \subseteq \mathbb{Z}$. Abbreviate $\varphi_{\mathbb{Z}} =: \varphi$, so this is now a factor map

$$(X, \mu, T) \rightarrow (A^{\mathbb{Z}}, \varphi_*\mu, S).$$

The entropy of the process $(\mathbf{X}, \mathcal{P})$ may be understood as the entropy rate of $\varphi_*\mu$, regarded as the law of a stationary sequence of A -valued random variables.

Having fixed \mathcal{P} , A and φ , the map $\varphi_{(-\infty;0)} : X \rightarrow A^{(-\infty;0)}$ is referred to as the **past** of the process $(\mathbf{X}, \mathcal{P})$. The measure μ may be disintegrated over $\varphi_{(-\infty;0)}$, giving a probability kernel

$$A^{(-\infty;0)} \rightarrow \text{Pr } X : z \mapsto \mu_z;$$

this is referred to as **conditioning on the past**. Various entropy-theoretic properties may be expressed in terms of these conditional measures: in the first place,

$$h(\mathbf{X}, \mathcal{P}) = \int H_{\mu_{\varphi_{(-\infty;0)}(x)}}(\mathcal{P}) \mu(dx) = \int H(\varphi_*\mu_{\varphi_{(-\infty;0)}(x)}) \mu(dx),$$

the expected Shannon entropy of \mathcal{P} given the past (see [Shi96, Subsection I.6.b]).

3.2 Compact models

Instead of finite partitions, much of our later work will rely on endowing X with a compact metric for which T is continuous. This is always possible by the following classical result (see, for instance, [Var85, Theorem 5.7]):

Theorem 3.4. *If (X, μ, T) is any jointly measurable p.-p. action of an l.c.s.c. group on a standard Borel probability space, then it is isomorphic as such to a jointly continuous action on a compact metric space with an invariant probability measure.* \square

In case (X, d^X) is a compact metric space, T is a jointly continuous action of \mathbb{Z} or \mathbb{R} on X , and $\mu \in \text{Pr}^T X$, we shall refer to (X, d^X, μ, T) as a **compact model p.-p. system** or **flow**. We shall work with compact models of our systems in much of the sequel. Of course, after choosing compact models, we must still allow arbitrary Borel (not necessarily continuous) factor maps between them. The key to using the metric space structure, in spite of this flexibility, will be Lusin's Theorem.

One can use such a choice of metric d^X to express the Kolmogorov-Sinai entropy. This relationship can be traced back to Feldman's work in [Fel80], and it is

worked out in detail (for actions of general unimodular amenable groups) by Ornstein and Weiss in [OW87, Part II]. We quickly recall some of the results that we need here, largely referring to that latter work.

First, for any compact model p.-p. system (X, d^X, μ, T) and any finite $F \subseteq \mathbb{Z}$, let

$$d_F^{\mathbf{X}}(x, x') := \sum_{n \in F} d^X(T^n x, T^n x').$$

This is a sequence of metrics on X . In terms of this construction, for any $r > 0$, one defines the **spatial r -entropy** $h(\mu, T, d^X, r)$ by

$$h(\mu, T, d^X, r) := \sup_{\varepsilon > 0} \liminf_{N \rightarrow \infty} \frac{1}{N} \text{cov}_{1-\varepsilon}((X, d_{[0;N]}^{\mathbf{X}}, \mu), rN). \quad (5)$$

Similarly, if (X, d^X, μ, T) is a compact p.-p. flow and $F \subseteq \mathbb{R}$ is measurable with finite measure, then

$$d_F^{\mathbf{X}}(x, x') := \int_F d^X(T^t x, T^t x') dt,$$

and the **spatial r -entropy** $h(\mu, T, d^X, r)$ is again given by (5), where now N is allowed to run through real values.

The connection between these spatial entropies and the Kolmogorov-Sinai entropy is the following, established in [Fel80, OW87]:

Proposition 3.5. *In the setting of either a compact model system or compact model flow, one has*

$$\sup_{r > 0} h(\mu, T, d^X, r) = \lim_{r \rightarrow 0} h(\mu, T, d^X, r) = h(\mu, T).$$

□

Corresponding to this, one would expect a relative of the Shannon-McMillan Theorem 3.1 for the exponential order of the μ -measure of a typical small-radius ball in the space $(X, d_{[0;N]}^{\mathbf{X}}, \mu)$, once N is large. Such a result is proved in [OW87, Section II.4, Theorem 5]. The related result that we will use below is actually a step on the way to their proof of that theorem.

Proposition 3.6 ([OW87, Section II.4, Proposition 3]). *For any $\beta \in (0, 1]$, $r > 0$ and $h^* < h(\mu, T, d^X, r)$, one has*

$$\text{cov}_{\beta}((X, d_{[0;N]}^{\mathbf{X}}, \mu), rN) > \exp(h^* N)$$

for all sufficiently large N .

□

The approach to entropy theory using compact metrics, rather than partitions, will be highly convenient in the rest of this paper. In Section 5, a new invariant of systems will be defined explicitly in terms of the sequences of metrics $d_{[0;N]}^{\mathbf{X}}$, and we will see that this ‘geometric’ definition leads naturally to a proof of isomorphism-invariance similar to a proof of the Kolmogorov-Sinai Theorem in terms of these metrics.

For the entropy theory of \mathbb{R} -actions, it has long been known that the metric-based approach is considerably cleaner and more efficient: this realization goes back to Feldman [Fel80], and stimulated the use of compact metrics in ergodic theory more generally. This program has recently been actively promoted by Vershik and his co-workers ([Ver10, VZP13]). As will become clear in Section 5, the present paper owes a great deal to this point of view.

Given a topological flow $T : \mathbb{R} \curvearrowright X$ with metric d^X , another dynamically-defined sequence of metrics on X may be obtained by supremizing over time-intervals, rather than integrating: for any nonempty compact $F \subseteq \mathbb{R}$, let

$$d_F^{\mathbf{X},\infty}(x, x') := \sup_{t \in F} d^X(T^t x, T^t x').$$

When it is necessary to distinguish this from the earlier metric, we will refer to the metrics $d_F^{\mathbf{X}}$ as **Hamming-like metrics** and to the metrics $d_F^{\mathbf{X},\infty}$ as **Bowen-Dinaburg metrics**. In topological dynamics, the asymptotic packing or covering numbers of the metrics $d_F^{\mathbf{X},\infty}$ are the basis of the Bowen-Dinaburg approach to topological entropy, but are not so directly related to Kolmogorov-Sinai entropy. However, it will be convenient to know later that given a topological flow (X, T) and an ergodic invariant probability μ , these alternative metrics may also be used to define $h(\mu, T)$. This has previously been proved in [Kat80a, Theorem 1.1]. However, we will need a slightly stronger, local version of that control, so we include a precise statement and proof here. Clearly $d_F^{\mathbf{X},\infty} \geq d_F^{\mathbf{X}}$, but we will need a result in the reverse direction.

Lemma 3.7. *If (X, d^X, μ, T) is an ergodic compact model flow, then for every $\varepsilon, \delta > 0$ there is a $\delta_1 > 0$ such that, for every $x \in X$ and $K \in \text{Int}(\mathbb{R})$ with $\mathcal{L}^1(K) \geq 1$, one has*

$$\text{cov}((B_{\delta_1 \mathcal{L}^1(K)}^{d_K^{\mathbf{X}}}(x), d_K^{\mathbf{X},\infty}), \delta) < \exp(\varepsilon \mathcal{L}^1(K)).$$

Proof. Clearly it suffices to prove this with $I = [0, a]$ for some $a \geq 1$. Let $N := \lfloor a \rfloor$, and observe that $N \geq a/2$.

By the joint continuity of T , there is some $\delta' > 0$ such that

$$\forall x, x' \in X, \quad d^X(x, x') < \delta' \implies \max_{t \in [-2, 2]} d^X(T^t x, T^t x') < \delta/2,$$

and now there is also some $\delta'' > 0$ such that

$$\forall x, x' \in X, \quad d^X(x, x') < \delta'' \implies \max_{t \in [-2, 2]} d^X(T^t x, T^t x') < \delta'.$$

This latter condition implies that if $d^X(T^n x, T^n x') \geq \delta'$ for some $n \in \mathbb{Z}$, then $d^X(T^t x, T^t x') \geq \delta''$ for all $t \in [n, n+1]$, and therefore

$$\int_0^a d^X(T^t x, T^t x') dt \geq \delta'' |\{n \in [0; N) \mid d^X(T^n x, T^n x') \geq \delta'\}|. \quad (6)$$

Let $\mathcal{P} = (P_1, \dots, P_m)$ be a Borel partition of X into sets of diameter less than δ' . Having chosen this, let $\eta \in (0, \varepsilon)$ be so small that in the space $[0; m]^N$ the cardinality of a Hamming ball of radius ηN is less than $e^{\varepsilon N}$ for all $N \geq 1$. Finally, choose $\delta_1 := \eta \delta'' / 2$.

After these preliminaries, suppose that $x, x' \in X$ satisfy $d_{[0, a]}^{\mathbf{X}}(x, x') < \delta_1 a \leq \eta \delta'' N$. Then (6) implies that

$$|\{n \in [0; N) \mid d^X(T^n x, T^n x') \geq \delta'\}| < \eta N.$$

Fix x , and for each $n \in \mathbb{Z}$ let

$$P_{n,0} := T^{-n}(B_{\delta'}^{d^X}(T^n x)) \quad \text{and} \quad P_{n,i} := T^{-n}(P_i) \quad \text{for } i = 1, 2, \dots, m.$$

Then the above estimates imply that

$$B_{\delta_1 a}^{d_{[0, a]}^{\mathbf{X}}}(x) \subseteq \bigcup_{\substack{(w_0, \dots, w_{N-1}) \in [0; m]^N \\ |\{n \in [0; N) \mid w_n \neq 0\}| < \eta N}} P_{0, w_0} \cap P_{1, w_1} \cap \dots \cap P_{N-1, w_{N-1}}.$$

By the choice of δ' , each individual intersection on the right here has $d_{[0, a]}^{\mathbf{X}, \infty}$ -diameter less than δ , and by the choice of η the number of such intersections appearing in this union is less than $e^{\varepsilon N}$. \square

3.3 Gibbs measures on mixing SFTs

The source of base systems for the examples in Theorem A is the class of Gibbs measures on mixing SFTs, and other invariant states on topological dynamical systems that can be suitably coded from these. These form the basic setting of the ‘thermodynamic formalism’. The standard monographs [Bow08, PP90] provide a good reference for most of our needs, and [Rue04, Sin72] largely cover the same material.

Given a finite alphabet A , we shall usually consider $A^{\mathbb{Z}}$ endowed with the metric

$$d(a, a') := \sum_{n \in \mathbb{Z}} 2^{-|n|} 1_{\{a_n \neq a'_n\}}.$$

We also endow $A^{(-\infty; 0]}$ with the analogous metric.

A function $A^{\mathbb{Z}} \rightarrow \mathbb{R}$ is **Hölder** if it is so with respect to d for some positive Hölder exponent, and similarly for a function $A^{(-\infty; 0]} \rightarrow \mathbb{R}$. A function on $A^{\mathbb{Z}}$ is **one-sided** if it factorizes through the coordinate projection $A^{\mathbb{Z}} \rightarrow A^{(-\infty; 0]}$. Motivated by the thermodynamic formalism, we will sometimes refer to a Hölder function restricted to any closed subset of $A^{\mathbb{Z}}$ as a **potential** (ignoring the many more general potentials that can be considered in the thermodynamic formalism).

As usual, a **subshift of finite type** ('SFT') in $A^{\mathbb{Z}}$ is a closed S -invariant subset $Y \subseteq A^{\mathbb{Z}}$ defined by a finite set of forbidden subwords. We always endow such an SFT with the restriction d^Y of the metric d above.

Given a topologically mixing SFT $Y \subseteq A^{\mathbb{Z}}$ and a potential $\varphi : Y \rightarrow \mathbb{R}$, there is always an associated **Gibbs measure** $\nu \in \text{Pr}^S Y$, uniquely characterized by the property that there are $c_1, c_2 \in (0, \infty)$ and $P \in \mathbb{R}$ such that

$$c_1 \exp \left(P|I| + \sum_{n \in I} \varphi(S^n y) \right) \leq \nu(\mathcal{P}^I(y)) \leq c_2 \exp \left(P|I| + \sum_{n \in I} \varphi(S^n y) \right) \quad (7)$$

for all $y \in Y$ and bounded discrete intervals $I \subseteq \mathbb{Z}$: see [Bow08, Theorem 1.4] or [PP90, Chapter 3]. Henceforth we shall refer to a triple (Y, ν, S) in which (Y, S) is a mixing SFT and ν is the Gibbs measure associated to some potential as a **mixing Gibbs system**.

Now let $\alpha : Y \rightarrow A$ be the time-zero coordinate map and let \mathcal{P} be the partition it generates. Let $Y^- := \alpha^{(-\infty; 0]}(Y) \subseteq A^{(-\infty; 0]}$. The Gibbs measure ν associated to a potential φ is constructed via its image $\nu^- := \alpha_*^{(-\infty; 0]} \nu \in \text{Pr} Y^-$. This image determines ν uniquely, by S -invariance. As in the proof of Ruelle's Perron-Frobenius Theorem (see [PP90, Theorem 2.2]), one may always find another Hölder function $\psi : Y^- \rightarrow \mathbb{R}$ such that

- $\varphi - \psi \circ \alpha^{(-\infty; 0]}$ is cohomologous to a constant over S among Hölder functions, and
- the Perron-Frobenius operator $C(Y^-) \rightarrow C(Y^-)$ defined by

$$L_\psi f(y) := \sum_{a \in A \mid ya \in Y^-} e^{\psi(ya)} f(ya)$$

satisfies $L_\psi 1_{Y^-} = 1_{Y^-}$, and otherwise has spectrum contained in a disk of radius strictly less than 1 (in this case the Perron-Frobenius operator is said to be 'normalized' [PP90, Chapter 2]).

Having found this ψ , the measure ν^- is the unique probability measure for which $L_\psi^* \nu^- = \nu^-$.

After reconstructing ν from ν^- , this Perron-Frobenius operator has the interpretation that for any bounded measurable function $f : Y^- \rightarrow \mathbb{R}$ and $r \in \mathbb{N}$ one has

$$\mathbb{E}_\nu(f \circ \alpha^{(-\infty;0]} \circ S^r \mid \mathcal{P}^{(-\infty;0]}) = (L_\psi^r f) \circ \alpha^{(-\infty;0]}.$$

In particular, if $y \mapsto \nu_y$ is the disintegration of ν over the strict past $\alpha^{(-\infty;0)} : Y \rightarrow A^{(-\infty;0)}$, then the equation $L_\psi^* \nu^- = \nu^-$ implies

$$\alpha_* \nu_y = \sum_{a \in A \mid ya \in Y^-} e^{\psi(ya)} \delta_a.$$

By the Hölder condition and the fact that Y is an SFT, there are $b < \infty$, $\beta \in (0, 1)$ and $N_0 \in \mathbb{N}$ such that

$$\begin{aligned} N \geq N_0 \quad \text{and} \quad y, y' \in Y^- \text{ with } \mathcal{P}^{[-N_0;0]}(y) = \mathcal{P}^{[-N_0;0]}(y') \\ \implies \{a \mid ya \in Y^-\} = \{a \mid y'a \in Y^-\} \text{ and } \max_{a \mid ya \in Y^-} |\varphi(ya) - \varphi(y'a)| < b\beta^N. \end{aligned}$$

This has proved the following.

Lemma 3.8 (Hölder continuity of conditional measures). *In the setting above there are $N_0 \in \mathbb{N}$, $b < \infty$ and $\beta \in (0, 1)$ such that for any $N \geq N_0$ one has*

$$\begin{aligned} y, y' \in Y^- \text{ with } \mathcal{P}^{[-N;0]}(y) = \mathcal{P}^{[-N;0]}(y') \\ \implies \alpha_* \nu_y \sim \alpha_* \nu_{y'} \quad \text{and} \quad e^{-b\beta^N} < \frac{d(\alpha_* \nu_y)}{d(\alpha_* \nu_{y'})} < e^{b\beta^N}. \end{aligned}$$

□

Corollary 3.9. *If (Y, ν, S) and \mathcal{P} are as above and also $p \in \mathbb{N} \cup \{0\}$, then*

$$\sup_{N \geq 1} \mathbb{I}_\nu(\mathcal{P}^{[-p;N+p]}; \mathcal{P}^{[-N-p;p]}) < \infty.$$

Proof. Suppose first that $p = 0$. The Chain Rule for mutual information ([CT06, Theorem 2.5.2]) gives

$$\mathbb{I}_\nu(\mathcal{P}^{[0;N]}; \mathcal{P}^{[-N;0]}) = \sum_{n=0}^{N-1} \mathbb{I}_\nu(S^{-n}(\mathcal{P}); \mathcal{P}^{[-N;0]} \mid \mathcal{P}^{[0;n]}). \quad (8)$$

Letting N_0 be as in Lemma 3.8, we obtain

$$e^{-b\beta^n} < \frac{d((\alpha \circ S^n)_*(\nu|_{\mathcal{P}^{[-N;n]}(y)}))}{d((\alpha \circ S^n)_*(\nu|_{\mathcal{P}^{[0;n]}(y)}))} < e^{b\beta^n} \quad \forall n \geq N_0,$$

and hence

$$D_{\text{KL}}((\alpha \circ S^n)_*(\nu|_{\mathcal{P}^{[-N;n]}(y)}) \mid (\alpha \circ S^n)_*(\nu|_{\mathcal{P}^{[0;n]}(y)})) < b\beta^n.$$

Integrating over $\nu|_{\mathcal{P}^{[0;n]}(y)}$ and recalling Equation (2), this gives

$$I_\nu(S^{-n}(\mathcal{P}); \mathcal{P}^{[-N;0]} \mid \mathcal{P}^{[0;n]}) < b\beta^n \quad \forall n \geq N_0.$$

Therefore the right-hand sum in (8) is bounded by

$$\sum_{n=0}^{N_0-1} I_\nu(S^{-n}(\mathcal{P}); \mathcal{P}^{[-N;0]} \mid \mathcal{P}^{[0;n]}) + \sum_{n=N_0}^{N-1} b\beta^n,$$

which remains bounded as $N \rightarrow \infty$ because $\sum_n b\beta^n$ is a convergent series.

Finally, if $p \geq 1$, then the definition and standard properties of mutual information give

$$\begin{aligned} & I_\nu(\mathcal{P}^{[-p;N+p]}; \mathcal{P}^{[-N-p;p]}) \\ &= H_\nu(\mathcal{P}^{[-N-p;p]}) + H_\nu(\mathcal{P}^{[-p;N+p]}) - H_\nu(\mathcal{P}^{[-N-p,N+p]}) \\ &\leq H_\nu(\mathcal{P}^{[-N-p;-N]}) + H_\nu(\mathcal{P}^{[0;p]}) + H_\nu(\mathcal{P}^{[-p;0]}) + H_\nu(\mathcal{P}^{[N;N+p]}) \\ &\quad + H_\nu(\mathcal{P}^{[-N;0]}) + H_\nu(\mathcal{P}^{[-N;0]}) - H_\nu(\mathcal{P}^{[-N,N]}) \\ &\leq 4pH_\nu(\mathcal{P}) + I_\nu(\mathcal{P}^{[0;N]}; \mathcal{P}^{[-N;0]}). \end{aligned}$$

□

3.4 Hölder cocycles over mixing SFTs

The structure of a generalized RWRS system seems to depend rather delicately on the cocycle σ which defines it. This subsection is given to various properties of such cocycles that will be needed later. The general flavour is of comparing them over long time-scales with Brownian motion. Such probabilistic limit theorems are a very classical subject in dynamics. They are all widely-known for simple random walk itself: a suitable reference is [R90]. In our slightly more general setting, much of what we need will be taken from Guivarc'h and Hardy's classic work [GH88], which in turn built on older methods of Nagaev for certain Markov chains [Nag57],

among others. A suitable invariance principle is proved by Bunimovich and Sinai in [BS81], or can be deduced from the strong invariance principles given in [PS75].

Let (Y, ν, S) be as in the previous subsection. A **cocycle** over a (Y, ν, S) will be a real-valued measurable function $\sigma : Y \rightarrow \mathbb{R}$. The term ‘cocycle’ will refer either to this function itself, or to the resulting function $\mathbb{Z} \times Y \rightarrow \mathbb{R}$ defined by the partial sums:

$$(n, y) \mapsto \begin{cases} \sum_{m=0}^{n-1} \sigma(S^m y) & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \\ -\sum_{m=n}^{-1} \sigma(S^m y) & \text{if } n \leq -1. \end{cases}$$

It will often be convenient to denote this value by σ_n^y . We may also think of it as a random function

$$\mathbb{Z} \rightarrow \mathbb{R} : n \mapsto \sigma_n^y$$

defined on the probability space (Y, ν) . For a fixed choice of y , we will refer to the function σ^y as a **cocycle-trajectory** to emphasize this point of view.

Mean-zero, Hölder cocycles enjoy a (weak) version of Donsker’s Invariance Principle. It is proved for any dynamical system admitting a suitable Markov partition (including our mixing Gibbs systems (Y, ν, S)) in [BS81]: see their Theorems 2” and 3. To formulate it, it will be convenient to introduce the maps $\text{traj}_N : \mathbb{R}^{\mathbb{Z}} \rightarrow C[0, 1]$ defined by

$$\text{traj}_N(\sigma)(t) := N^{-1/2}((Nt - \lfloor Nt \rfloor)\sigma_{\lfloor Nt \rfloor} + (\lfloor Nt + 1 \rfloor - Nt)\sigma_{\lfloor Nt \rfloor})$$

(that is, traj_N rescales σ horizontally by N^{-1} and vertically by $N^{-1/2}$, and then interpolates linearly to produce a function on $[0, 1]$). Similarly, define $\text{traj}_{-N} : \mathbb{R}^{\mathbb{Z}} \rightarrow C[0, 1]$ by

$$\text{traj}_{-N}(\sigma)(s) := N^{-1/2}((-Ns - \lfloor -Ns \rfloor)\sigma_{\lfloor -Ns \rfloor} + (\lfloor -Ns + 1 \rfloor + Ns)\sigma_{\lfloor -Ns \rfloor}).$$

Theorem 3.10 (Invariance principle). *If (Y, σ) is a mixing Gibbs system and $\sigma : Y \rightarrow \mathbb{R}$ is a Hölder cocycle with $\int \sigma d\nu = 0$, then there is some $c \geq 0$ for which the **Invariance Principle** holds:*

$$\text{traj}_N(\sigma^y) \xrightarrow{\text{law}} cB \quad \text{as } N \rightarrow \infty,$$

where the left-hand side is regarded as a random variable on the probability space (Y, ν) , and the right hand side has law $\mathbb{W}_{[0,1]}$. Moreover, $c = 0$ if and only if σ is a coboundary over S among Hölder functions. \square

In view of its rôle in the above theorem, we shall call the constant c^2 the **effective variance** of (\mathbf{Y}, σ) . Henceforth we will work with non-coboundaries, and will generally normalize so that the effective variance is 1.

Definition 3.11. A *well-distributed pair* is a pair (\mathbf{Y}, σ) in which $\mathbf{Y} = (Y, \nu, S)$ is a mixing Gibbs system and $\sigma : Y \rightarrow \mathbb{R}$ is a Hölder non-coboundary with $\int \sigma d\nu = 0$ and with effective variance 1.

The next result gives the optimum rate of convergence to a Gaussian law for the distribution of σ_N^y for a fixed N .

Theorem 3.12 (Berry-Esseen property: [GH88, Théorème B.IV.2]). *If (\mathbf{Y}, σ) is well-distributed, then*

$$\sup_{t \in \mathbb{R}} |\nu\{\sigma_N^y \leq t\sqrt{N}\} - N(-\infty, t)| \lesssim_{\mathbf{Y}, \sigma} \frac{1}{\sqrt{N}} \quad \forall N \geq 1, \quad (9)$$

where $N(-\infty, t)$ is the cumulative distribution function of a standard Gaussian. \square

3.5 The Enhanced Invariance Principle

In addition to the preceding results on cocycle-distribution, we will also need an enhancement of the conclusion of Theorem 3.10 which simultaneously describes the frequency of visits of σ^y to different regions in \mathbb{R} . To formulate this, given $y \in Y$ and a nonempty finite subset $F \subseteq \mathbb{Z}$, let

$$\gamma_F^y := \frac{1}{|F|} \sum_{n \in F} \delta_{\sigma_n^y}.$$

This is the **occupation measure of σ over the set of times F** .

Let $L^B(u)$, $u \in \mathbb{R}$, be Brownian local time at time 1, regarded as a $C_c(\mathbb{R})$ -valued random variable on the space $(C_0(0, 1], W_{[0, 1]})$ (see, for instance, [Kal02, Chapter 22]). Observe that if $\varphi : \mathbb{R} \rightarrow [0, \infty)$ is a mollifier and θ is a Radon measure on \mathbb{R} , then the convolution $\varphi \star \theta$ may always be understood as the smooth function

$$u \mapsto \int \varphi(v - u) \theta(dv).$$

The following is the additional property of a well-distributed pair that we will need.

Definition 3.13 (Enhanced Invariance Principle). *The well-distributed pair (\mathbf{Y}, σ) satisfies the **Enhanced Invariance Principle** if there is a mollifier φ such that*

$$(\text{traj}_N(\sigma^y), ((\varphi \star \gamma_{[0; N]}^y)(\sqrt{N}u))_{u \in \mathbb{R}}) \xrightarrow{\text{law}} (B, L^B)$$

for the product of the uniform and locally uniform topologies on $C_0(0, 1] \times C_c(\mathbb{R})$. As before, the left- and right-hand sides here are understood as random variables on (Y, ν) and $(C_0(0, 1], W_{[0,1]})$, respectively.

I strongly suspect that every well-distributed pair satisfies the Enhanced Invariance Principle, so that the above could instead be introduced as a theorem. If σ is aperiodic (see [GH88]), then the above convergence should actually hold for every mollifier φ . If σ is cohomologous to an $\ell\mathbb{Z}$ -valued cocycle for some $\ell > 0$, say $\sigma = \tau + f \circ S - f$, then the occupation measures of σ^y are ‘adjustments’ of those of τ^y , which are supported on $\ell\mathbb{Z}$. However, the above should still hold provided φ is strictly positive on an interval $[-a, a]$ with $a > \max\{\ell, \|f\|_\infty\}$. The proofs of these results should be based on the same spectral analysis of the complex Ruelle operator as in [GH88] or [PP90, Chapter 4]. However, as far as I know this result has appeared in the literature only in special cases:

- In case (Y, σ) is the pair of a simple random walk, then it follows from a much stronger classical coupling result between occupation measures of simple random walk and Brownian motion ([R90, Theorem 10.1]).
- The generalization to partial sums of Markov chains was recently established by Bromberg and Kosloff [BK14], building on older results of Borodin [Bor81].

Thus, our Theorem A is unconditional in either of the above cases. The first of these covers the classical RWRSs.

I understand that the full generalization (even to the still-broader setting of finite-variance Hölder cocycles on Gibbs-Markov shifts — see [Aar97, Chapter 4], [AD01]) will be the subject of future work by Bromberg.

Similar results for cocycles over general Young towers appear as [DSV08, Theorem 9] and [NS12, Proposition 3], but focusing only on finite-dimensional marginals.

The Enhanced Invariance Principle will be used to prove Theorem 5.23, which evaluates our forthcoming new invariant in the case of generalized RWRS systems. In fact, it will be needed only for proving the lower-bound half of that Theorem, in Sections 8 and 9.

Remark 3.14. *In recent years there has been considerable interest in generalizing probabilistic limit theorems for ergodic sums to dynamical systems that admit a more general Markov-Gibbs structure or a suitable Young tower ([You98]): see, for instance, [AD01, Gou05, SV04, DSV08, Xia09] and the many further references there. A fairly gentle introduction to the use of Young towers is in [Bal00, Chapter 4], and related material can also be found in the monograph [HH01].*

In suspect that Theorem A can be extended to the study of generalized RWRS systems with base and cocycle given by one of these more general settings. However, in addition to the Enhanced Invariance Principle, one would need some restriction on the relevant generating partition to obtain an analog of Corollary 3.9. \triangleleft

Remark 3.15. Rudolph's work in [Rud88] studies systems satisfying a rather different kind of convergence to Brownian motion: his **asymptotically Brownian** cocycles σ admit some $\eta > 0$ and a (ν, W) -coupling P such that for P -a.e. (y, B) one has

$$|\sigma_n^y - B_n| = o(n^{1/2-\eta}) \quad \text{as } n \longrightarrow \infty.$$

This definition follows Philipp and Stout [PS75], who establish that a wide variety of examples are η -asymptotically Brownian for some η . In principle, the existence of such a coupling is significantly stronger than the conclusion of Theorem 3.10, but it also does not seem to imply the Enhanced Invariance Principle without some additional arguments as in [BK14], so our assumptions on σ are actually somewhat askew to Rudolph's. It could be that our Theorem A gives new examples of non-Bernoulli K -automorphisms, not covered by [Rud88], but I do not know of any specific systems that fall into this gap. \triangleleft

4 Informal discussion of the RWRS marginal metrics

This section is discursive. It is not needed for the logic in the rest of the paper, but offers some motivation for the constructions that follow.

The new invariant below is defined in terms of the ‘marginal’ m.p. spaces that arise from a given compact model of a generalized RWRS system. This section will begin with a sketch of the ‘marginal’ m.p. spaces that arise from the canonical generating partition of a classical RWRS example.

4.1 Conditioning on the scenery, or the past

Let $\alpha : \{\pm 1\}^{\mathbb{Z}} \times C^{\mathbb{Z}} \longrightarrow \{\pm 1\} \times C$ be the obvious generating partition for RWRS_μ , and let $\rho_N = \alpha_*^{[0;N)} \rho$ be the distribution of the (α, N) -name, as in Subsection 1.3. Given a scenery distribution $\mu \in \text{Pr}^S C^{\mathbb{Z}}$, let μ_I be its marginal on C^I for any $I \subseteq \mathbb{Z}$.

Let d_α be the pseudometric on $\{\pm 1\}^{\mathbb{Z}} \times C^{\mathbb{Z}}$ given by the pullback under α of the complete metric on $\{\pm 1\} \times C$. Then the marginal psm.p. spaces given by $(d_\alpha)_{[0;N)}^{\text{RWRS}_\mu}$ are likewise pulled back from the finite m.p. spaces

$$((\{\pm 1\} \times C)^N, d_{\text{Ham}}, \rho_N).$$

We now sketch a provisional description of these m.p. spaces. This is in terms of the 1-Lipschitz quotient map

$$\begin{array}{c} ((\{\pm 1\} \times C)^N, d_{\text{Ham}}, \rho_N) \\ \downarrow \\ (\{\pm 1\}^N, d_{\text{Ham}}, \nu_{1/2}^{\otimes N}). \end{array}$$

The idea is to describe ρ_N as a lift of $\nu_{1/2}^{\otimes N}$ through this map.

Consider a fixed scenery $c = (c_m)_m \in C^{\mathbb{Z}}$, and define the function $F_c : \{\pm 1\}^N \rightarrow C^N$ by

$$F_c((y_n)_{n=0}^{N-1}) = (c_{\sigma_0^y}, c_{\sigma_1^y}, \dots, c_{\sigma_{N-1}^y}).$$

Clearly this output depends only on the finite portion $c|_{\sigma_{[0;N)}^y}$ of c . Let

$$\rho_{N,c} := \int_{\{\pm 1\}^N} \delta_{(y, F_c(y))} \nu_{1/2}^{\otimes N}(\mathrm{d}y),$$

the result of lifting $\nu_{1/2}^{\otimes N}$ to the graph of F_c .

We can now write the lifted measure ρ_N as the average of the conditional measures of ρ_N given the scenery, and these latter are precisely the graph-supported measures $\rho_{N,c}$:

$$\rho_N = \int_{C^{\mathbb{Z}}} \rho_{N,c} \mu(\mathrm{d}c). \quad (10)$$

This decomposition of ρ_N is obtained canonically from the process $(\text{RWRS}_\mu, \alpha)$: it is the pushforward under $\alpha^{[0;N)}$ of the disintegration of ρ over the strict past $\alpha^{(-\infty;0)}$. This is because

- on the one hand, the past of the simple random walk is independent of the future,
- but on the other, simple random walk is recurrent, so the past of the whole process a.s. determines the scenery exactly.

4.2 Separating the conditional measures

We can now describe the overall strategy of the proof of non-Bernoullicity in [Kal82]. The heart of Kalikow's work is to prove that there are arbitrarily large N for which the following holds.

Theorem 4.1 ([Kal82]). *For a fixed sequence of walk-steps $y \in \{\pm 1\}^N$ and a fixed scenery $c' \in C^{\mathbb{Z}}$, it holds for most $c \in C^{\mathbb{Z}}$ that*

$$d_{\text{Ham}}((y, F_c(y)), \text{spt } \rho_{N,c'}) > 10^{-20} N,$$

where ‘most’ means ‘with high probability as $N \rightarrow \infty$ ’.

□

(Indeed, Kalikow actually proves this with Feldman’s weaker \bar{f} -metric in place of d_{Ham} .)

By Fubini’s Theorem, the above implies that for $(\mu \otimes \mu)$ -most pairs (c, c') there is a subset $W_{c,c'} \subseteq \{\pm 1\}^N$ such that

$$\nu_{1/2}^{\otimes N}(W_{c,c'}) = 1 - o(1) \quad \text{and} \quad d_{\text{Ham}}((\text{id}, F_c)(W_{c,c'}), \text{spt } \rho_{N,c'}) > 10^{-20} N.$$

This implies that a typical pair of conditional measures $\rho_{N,c}, \rho_{N,c'}$ are $\Omega(N)$ -separated in the Wasserstein metric associated to d_{Ham} , and hence that RWRS_μ does not satisfy the Very Weak Bernoulli condition.

An alternative description of this reasoning, more intrinsic to the metric geometry of $(\{\pm 1\} \times C)^N, d_{\text{Ham}}, \rho_N$, uses a different characterization of Bernoullicity in terms of measure concentration.

Definition 4.2 ((Almost) Exponential measure concentration). *Let (X_n, d_n, μ_n) be a sequence of compact psm.p. spaces. The sequence exhibits **exponential measure concentration** if for every $\delta > 0$ there is a $c > 0$ such that for any Borel set $U \subseteq X_n$ one has*

$$\mu_n(U) \geq e^{-cn} \implies \mu_n(B_\delta^{d_n}(U)) \geq 1 - e^{-cn}$$

for all sufficiently large n . The constant c is the **exponential rate** of this concentration at distance δ .

The sequence exhibits **almost exponential measure concentration** if there is a sequence of Borel subsets $X'_n \subseteq X_n$ such that $\mu_n(X'_n) \rightarrow 1$ and $(X'_n, d_n, (\mu_n)|_{X'_n})$ exhibits exponential measure concentration.

Theorem 4.3 (Exponential measure concentration in Bernoulli shifts). *Let $\mathbf{X} = (X, \mu, T)$ be a p.p. system of entropy $h < \infty$, and let \mathcal{P} be a finite generating partition of \mathbf{X} . Then \mathbf{X} is Bernoulli if and only if the sequence of psm.p. spaces $(X, N^{-1}d_{[0;N]}^{\mathcal{P}}, \mu)$ exhibits almost exponential measure concentration.* □

This is essentially the same as [Shi96, Theorem III.4.3], or can be quickly deduced from the implications proved in [KM10, Chapter 5]. It was introduced explicitly into ergodic theory by Marton and Shields in [MS94], where it was called

the ‘blowing-up property’. It is, however, also very close to Thouvenot’s notion of ‘extremality’, presented in [Tho02, Definition 6.3] but devised much earlier. These properties are now properly viewed as instances of the general phenomenon of concentration of measure: see, for instance, [Led01] or [Gro01, Chapter 3 $\frac{1}{2}$] for an introduction.

Returning to RWRS $_{\mu}$, now fix some very small $\varepsilon > 0$. Since simple random walk is diffusive, we may pick some large distance-cutoff $R \in \mathbb{N}$ so that the set

$$Y_N := \{y \in \{\pm 1\}^N \mid \sigma_{[0;N)}^y \subseteq [-R\sqrt{N}; R\sqrt{N}]\}$$

has $\nu_{1/2}^{\otimes N}(Y_N) > 1 - \varepsilon$ for all sufficiently large N . Let $Z_N := Y_N \times C^N$, so

$$\rho_N(Z_N) = \rho_{N,c}(Z_N) = \nu_{1/2}^{\otimes N}(Y_N) > 1 - \varepsilon,$$

because ρ_N and each $\rho_{N,c}$ is a lift of $\nu_{1/2}^{\otimes N}$.

In addition, Theorem 3.1 gives subsets $X_{I,\varepsilon}^{\text{SM}} \subseteq C^I$ for each bounded discrete interval $I \subseteq \mathbb{Z}$ such that

$$|X_{I,\varepsilon}^{\text{SM}}| \leq \exp((h(\mu, S) + \varepsilon)|I|) \quad \text{and} \quad \mu_I(X_{I,\varepsilon}^{\text{SM}}) > 1 - o(1) \text{ as } |I| \longrightarrow \infty.$$

Let $X_N := X_{[-R\sqrt{N}; R\sqrt{N}], \varepsilon}^{\text{SM}}$.

If $y \in Y_N$, then $F_c(y)$ depends only on the portion $c|_{[-R\sqrt{N}; R\sqrt{N}]}$, and therefore $(\rho_{N,c})|_{Z_N}$ depends only on $c|_{[-R\sqrt{N}; R\sqrt{N}]}$. With some slight abuse of notation, it follows that

$$\begin{aligned} \rho_N \approx_{\varepsilon} (\rho_N)|_{Z_N} &= \int_{C^{[-R\sqrt{N}; R\sqrt{N}]}} (\rho_{N,c})|_{Z_N} \mu_{[-R\sqrt{N}; R\sqrt{N}]}(dc) \\ &\approx_{\varepsilon} \int_{X_N} (\rho_{N,c})|_{Z_N} \mu_{[-R\sqrt{N}; R\sqrt{N}]}(dc) \end{aligned} \quad (11)$$

for sufficiently large N . Thus, most of the mass in the decomposition (10) is a convex combination of $|X_N| \leq \exp(2R(h(\mu, S) + \varepsilon)\sqrt{N})$ different measures supported on the graphs of the functions $F_c|_{Y_N}$.

Now, Kalikow’s conclusion in Theorem 4.1 may easily be adapted to see that most pairs of the measures in the coarsened decomposition (11) are also well-separated in the Wasserstein metric. Since there are only $\exp(O(\sqrt{N}))$ of these measures, an easy argument now shows that this precludes ρ_N from exhibiting almost exponential measure concentration.

4.3 Significance for approximate recovery of the scenery

Our work below will re-use the main ideas from Kalikow's proof of Theorem 4.1, but to a different end. As discussed in the Introduction, the scenery entropy $h(\mu, S)$ should appear in estimates on the mutual information $I_\rho(\alpha^{[-N;0]}; \alpha^{[0;N]})$. However, we need to make this quantity more robust, by asking after the information about a pair (y, c) that can be recovered if one knows the output strings $\alpha^{[-N;0]}(y, c)$ and $\alpha^{[0;N]}(y, c)$ only approximately.

We still expect this information to reside in that part of the scenery visited by both of the trajectories $\sigma_{[-N;0]}^y$ and $\sigma_{[0;N]}^y$, so the heart of the matter is now the ability to recover $c|_{\sigma_{[0;N]}^y}$ approximately if one only knows

$$\alpha^{[0;N]}(y, c) = (y, F_c(y)).$$

approximately.

This is difficult, because the map $(c, y) \mapsto F_c(y)$ can contract the relevant Hamming distances very greatly.

Example 4.4. *If $y = (y_n)_{n \in [0;N]}$ and $y' = (y'_n)_{n \in [0;N]}$ are chosen so that $y_0 = 1$, $y'_0 = -1$, but $y_n = y'_n$ for all $n \in [1; N]$, then*

$$d_{\text{Ham}}(y, y') = 1,$$

but

$$\sigma_n^y = \sigma_n^{y'} + 2 \quad \forall n \in [1; N].$$

Therefore, if $c \in C^{\mathbb{Z}}$ and $c' := S^2 c$, then $F_c(y)$ and $F_{c'}(y')$ agree in every coordinate in $[1; N]$. Thus

$$d_{\text{Ham}}((y, F_c(y)), (y', F_{c'}(y'))) = 1,$$

even though c and c' could be very far apart according to the relevant Hamming metric. More subtle examples of this phenomenon are described in [Lin99]. \triangleleft

Therefore, if one knows $y|_{[0;N]}$ only up to a small Hamming-metric error, it could happen that $c|_{\sigma_{[0;N]}^y}$ cannot be recovered up to a small Hamming error from the output-string $F_c(y)$. In order to work around this problem, we will need to set up a different, weaker sense in which approximate knowledge of $(y, F_c(y))$ constrains the possible choices of c , which is still strong enough that we obtain the same leading-order asymptotics as for true mutual information.

In view of the above example, a natural conjecture in this direction would be that, after excluding a small-probability set of ‘bad’ trajectories y , it holds that

$$d_{\text{Ham}}((y, F_c(y)), (y', F_{c'}(y'))) \approx 0 \\ \implies \frac{|\sigma_{[0;N]}^y \Delta \sigma_{[0;N]}^{y'}|}{\sqrt{N}} \approx 0 \quad \text{and} \quad \bar{f}_{\sigma_{[0;N]}^y}(c, c') \approx 0,$$

where \bar{f}_I is Feldman’s metric over a bounded discrete interval I from [Fel76]. Unfortunately, I do not know how to prove this. Instead, we will work with an even weaker (and significantly more complicated) notion of similarity between sceneries. Setting up this notion and then proving the analog of the above implication will be the most substantial part of our work, and will occupy most of Sections 8 and 9.

Remark 4.5. *The above discussion is suggestive of a link with the ‘scenery reconstruction problem’, which asks whether the entire scenery c can eventually be reconstructed from only the output string $(c_0, c_{y_0}, c_{\sigma_1^y}, \dots)$, with probability 1 in the choice of (y_0, y_1, \dots) . Much is known about that problem, but the methods do not seem well-adapted to the problem of ‘approximate reconstruction’ described above. Essentially, this is because in those works the scenery is reconstructed only very ‘slowly’: that is, the patch $c|_{[-m;m]}$ can be recovered with high probability only once one has seen $(c_0, c_{y_0}, c_{\sigma_1^y}, \dots, c_{\sigma_M^y})$ for some $M \gg m^2$. The best control on the necessary M is some high-degree polynomial in m , obtained by Matzinger and Rolles in [MR03]. They conjecture that it suffices to use $M \ll m^{2+\varepsilon}$ for any $\varepsilon > 0$, but this would still be too large for our purposes. Nevertheless, it would be interesting to know of any conceptual intersection between their methods and ours.*

More background on scenery reconstruction can be found in Section 3 of the survey [dHS06], and in the dedicated surveys [ML] and [Kes98]. \triangleleft

Remark 4.6. *Another proposal for an invariant of systems that should capture something like the above sequence of mutual informations is Vershik’s ‘secondary entropy’, formulated in [Ver00, Section 7]. Essentially, it amounts to quantifying the failure of the Very Weak Bernoulli property of an abstract process $(\mathbf{Z}, \mathcal{R})$ in terms of packings numbers within the space of future-name distributions. However, I am not aware that this quantity has been shown to be invariant under isomorphisms of processes, and I also do not see how to estimate it accurately enough for RWRS processes. Nevertheless, Vershik’s idea was a key motivation for the invariant that we define below.* \triangleleft

5 The new invariant

This section is largely concerned with general metric or pseudometric spaces, or general compact model p.-p. systems $\mathbf{X} = (X, d^X, \mu, T)$. For these systems, the key to our new invariant will be to consider not just the asymptotic behaviour of the sequence of metrics $d_{[0;N]}^{\mathbf{X}}$ on (X, μ) , but that of the sequence of *pairs* of metrics

$$d_{[-N;0]}^{\mathbf{X}} \quad \text{and} \quad d_{[0;N]}^{\mathbf{X}}.$$

5.1 Pair-metric spaces and bi-neighbourhoods

Definition 5.1. A *pair-metric space* is a triple (X, d_1, d_2) in which d_1 and d_2 are two compact metrics generating the same topology on X . A *pair-m.m.* (resp. *pair-m.p.*) *space* is a quadruple (X, d_1, d_2, μ) consisting of a pair-metric space and a finite Radon (resp. Radon probability) measure on X .

Note that we always assume compactness without mentioning it in the nomenclature. It will be important that d_1 and d_2 do not generate different topologies.

It will be convenient to allow also pairs of pseudometrics.

Definition 5.2. A *pair-pseudometric space* is a triple (X, d_1, d_2) in which X is a standard Borel space and d_1 and d_2 are two totally bounded Borel pseudometrics $X \times X \rightarrow [0, \infty)$. A *pair-psm.m.* (resp. *pair-psm.p.*) *space* is a quadruple (X, d_1, d_2, μ) consisting of a pair-psuedometric space and a finite Radon (resp. Radon probability) measure on X .

Note again that we always assume total boundedness without mentioning it in the nomenclature.

Definition 5.3. If (X, d_1^X, d_2^X) and (Y, d_1^Y, d_2^Y) are pair-pseudometric spaces and $c, L > 0$, then a map $f : X \rightarrow Y$ is *L-pair-Lipschitz* (resp. *c-almost L-pair-Lipschitz*) if it is *L-Lipschitz* (resp. *c-almost L-Lipschitz*) as a map $(X, d_i^X) \rightarrow (Y, d_i^Y)$ for $i = 1, 2$.

Example 5.4. Let $X = [0, 1]^3$, and let

$$d_1((x_1, x_2, x_3), (x'_1, x'_2, x'_3)) := |x_1 - x'_1| + |x_2 - x'_2|$$

and

$$d_2((x_1, x_2, x_3), (x'_1, x'_2, x'_3)) := |x_2 - x'_2| + |x_3 - x'_3|.$$

Then (X, d_1, d_2) is a pair-pseudometric space in which neither d_1 nor d_2 is a metric. \triangleleft

Given a p.-p. system (X, μ, T) and a totally bounded Borel pseudometric d^X on X , we will consider the sequence of pair-psm.p. spaces

$$(X, d_{[-N;0]}^X, d_{[0;N]}^X, \mu), \quad N \geq 1.$$

These are referred to as the **marginal pair-psm.p. spaces** of (X, d^X, μ, T) .

Our new invariant will involve some quantification of how much information is ‘robust’ under both of the pseudometrics $d_{[-N;0]}^X$ and $d_{[0;N]}^X$ on X . This will be made precise via the following notion.

Definition 5.5. *Let (X, d_1, d_2) be a pair-pseudometric space and $\delta \geq 0$. The δ -bi-neighbourhood in (X, d_1, d_2) around a point $x \in X$ is the set*

$$B_\delta^{d_2}(B_\delta^{d_1}(x)).$$

A pair of points $(x, y) \in X^2$ is δ -bi-separated in (X, d_1, d_2) if

$$B_\delta^{d_2}(B_\delta^{d_1}(x)) \cap B_\delta^{d_2}(B_\delta^{d_1}(y)) = \emptyset.$$

These definitions are not symmetrical in d_1 and d_2 ; though possibly disappointing, this will not matter in the sequel.

Give a subset $F \subseteq X$, its δ -bi-neighbourhood is

$$B_\delta^{d_2}(B_\delta^{d_1}(F)) = \bigcup_{x \in F} B_\delta^{d_2}(B_\delta^{d_1}(x)).$$

The property of bi-separation will not be used much below, but it gives some useful first intuition for bi-neighbourhoods. Explicitly, $x, y \in X$ are δ -bi-separated if for any $x', y', z \in X$, the following four inequalities cannot all hold:

$$d_1(x, x') \leq \delta, \quad d_2(x', z) \leq \delta, \quad d_2(z, y') \leq \delta \quad \text{and} \quad d_1(y', y) \leq \delta.$$

Thus, this asserts that one cannot move from x to y by taking a jump which is very small for the metric d_1 , then two jumps which are very small for d_2 , then another jump which is very small for d_1 .

Clearly if (x, y) is δ -bi-separated, then one must have $d_1(x, y) \geq 2\delta$ and also $d_2(x, y) \geq 2\delta$. However, the reverse of this implication need not hold, even approximately.

Example 5.6. *Recall the pair-pseudometric space in Example 5.4, and let $x = (x_1, x_2, x_3)$ and $x' = (x'_1, x'_2, x'_3)$ be points of $[0, 1]^3$. Then*

$$B_\delta^{d_2}(B_\delta^{d_1}(x)) := \{(x'_1, x'_2, x'_3) \mid |x_2 - x'_2| \leq 2\delta\},$$

and so x, y are δ -bi-separated if and only if

$$|x_2 - y_2| \geq 4\delta.$$

In particular, the points $(0, 1, 0)$ and $(1, 1, 1)$ are far apart according to both d_1 and d_2 , but are not δ -bi-separated for any $\delta > 0$. \triangleleft

The following is now the obvious analog of (3) for bi-neighbourhoods.

Definition 5.7. Let (X, d_1, d_2, μ) be a pair-psm.m. space. For $a, \delta > 0$, the *a-partial δ -bi-covering number* is

$$\text{bicov}_a((X, d_1, d_2, \mu), \delta) := \min\{|F| \mid F \subseteq X, \mu(B_\delta^{d_2}(B_\delta^{d_1}(F))) > a\}.$$

We also define simply

$$\text{bicov}((X, d_1, d_2), \delta) := \min\{|F| \mid F \subseteq X, B_\delta^{d_2}(B_\delta^{d_1}(F)) = X\},$$

by analogy with classical covering numbers.

Remark 5.8. Similarly, there is an obvious definition of $\text{bipack}((X, d_1, d_2), \delta)$ in terms of bi-separation. However, unlike for classical covering and packing numbers, I believe there are no simple relations between bicov and bipack . In essence, this is because the estimates relating covering and packing numbers rely on the inclusion

$$B_\delta(B_\delta(x)) \subseteq B_{2\delta}(x) \quad \forall x, \delta.$$

However, no corresponding inclusion need hold in the pair-pseudometric setting: given any $\delta \ll \delta'$, one can easily concoct examples in which $B_\delta^{d_2}(B_\delta^{d_1}(B_{\delta'}^{d_2}(B_{\delta'}^{d_1}(x))))$ is much larger than $B_{\delta'}^{d_2}(B_{\delta'}^{d_1}(x))$.

In fact, one could develop most of the rest of the present paper using bi-packing instead of bi-covering numbers, and I believe they would still serve to distinguish RWRS systems. Bi-covering numbers seem to require slightly simpler estimates, so we focus on them. However, it would be interesting to know of examples of systems for which these two different quantities give genuinely different invariants, perhaps with one behaving trivially and the other non-trivially. \triangleleft

Now suppose that (X, d_1^X, d_2^X) and (Y, d_1^Y, d_2^Y) are pair-pseudometric spaces, that $c, L > 0$, and that $\Phi : X \rightarrow Y$ is a c -almost L -pair-Lipschitz map. In this case, one has the obvious inclusion

$$\Phi(B_\delta^{d_2^X}(B_\delta^{d_1^X}(x))) \subseteq B_{L\delta+c}^{d_2^Y}(B_{L\delta+c}^{d_1^Y}(\Phi(x))) \quad \forall x \in X,$$

and hence also

$$\Phi(B_\delta^{d_2^X}(B_\delta^{d_1^X}(F))) \subseteq B_{L\delta+c}^{d_2^Y}(B_{L\delta+c}^{d_1^Y}(\Phi(F))) \quad \forall F \subseteq X. \quad (12)$$

This leads immediately to the following.

Lemma 5.9. *Let $c, L, a, \delta > 0$. Suppose that (X, d_1^X, d_2^X) and (Y, d_1^Y, d_2^Y) are pair-pseudometric spaces, that μ is a finite Borel measure on X , and that $\Phi : X \rightarrow Y$ is a c -almost L -pair-Lipschitz map. Then*

$$\text{bicov}_a((X, d_1^X, d_2^X, \mu), \delta) \geq \text{bicov}_a((Y, d_1^Y, d_2^Y, \Phi_*\mu), L\delta + c).$$

Proof. If $F \subseteq X$, then (12) implies

$$\Phi_*\mu(B_{L\delta+c}^{d_2^Y}(B_{L\delta+c}^{d_1^Y}(\Phi(F)))) \geq \Phi_*\mu(\Phi(B_\delta^{d_2^X}(B_\delta^{d_1^X}(F)))) \geq \mu(B_\delta^{d_2^X}(B_\delta^{d_1^X}(F))).$$

□

5.2 Passing to subsets

By analogy with (5), a natural place to look for a new invariant of a p.-p. system (X, μ, T) would be in the asymptotic behaviour of

$$\text{bicov}_{1-\varepsilon}((X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu), \delta N) \quad (13)$$

as $N \rightarrow \infty$ for a suitable choice of (pseudo)metric d on X , possibly then also sending $\varepsilon \downarrow 0$ and $\delta \downarrow 0$ in the right order.

The arguments below can easily be adapted to show that one does obtain isomorphism-invariants this way. However, as far as I know, they do not achieve the purpose of distinguishing RWRS systems. Instead, our new invariant will be obtained from the bi-covering numbers of various *subspaces* of $(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu)$.

The need to pass to subsets in a controlled way will be discussed in more detail shortly. There are surely many ways to do this which will lead to a more refined invariant. The procedure of this subsection is the simplest I have found to work, but is by no means canonical.

The key next point to emphasize is that bi-neighbourhoods can behave much more subtly than ordinary neighbourhoods under passing to subspaces. If $x \in Y \subseteq X$, then the δ -bi-neighbourhood of x in the pair-metric subspace (Y, d_1, d_2) is

$$Y \cap B_\delta^{d_2}(Y \cap B_\delta^{d_1}(x)), \quad (14)$$

and this may be much smaller than just $Y \cap B_\delta^{d_2}(B_\delta^{d_1}(x))$. Crucially, this means that bi-covering numbers can *increase* under passing to subsets.

Example 5.10. *Let (X, d_1, d_2) be as in Example 5.4, and let $U := \{(x, 0, x) \mid x \in [0, 1]\} \subset [0, 1]^3$. One has*

$$d_i((x, 0, x), (y, 0, y)) = |x - y| \quad \text{for both } i = 1, 2,$$

and so within the pair-pseudometric space (U, d_1, d_2) , the δ -bi-neighbourhood of $(x, 0, x)$ is precisely

$$\{(y, 0, y) \mid |x - y| \leq 2\delta\}.$$

By contrast, letting $V := [0, 1] \times \{0\} \times [0, 1]$, for any $(x, 0, x'), (y, 0, y') \in V$ one has

$$d_1((x, 0, x'), (x, 0, y')) = d_2((x, 0, y'), (y, 0, y')) = 0,$$

and so for every point of V , its δ -bi-neighbourhood in (V, d_1, d_2) is the whole of V , for any $\delta > 0$.

Therefore, even though $U \subseteq V$, we obtain

$$\text{bicov}((U, d_1, d_2), \delta) \sim (2\delta)^{-1} \quad \text{whereas} \quad \text{bicov}((V, d_1, d_2), \delta) = 1 \quad \forall \delta > 0.$$

◁

Now consider some further parameters $\alpha \in [1, \infty)$ and $\kappa > \kappa' > 0$.

Definition 5.11. For a pair-psm.p. space (X, d_1, d_2, μ) , $\alpha \in [1, \infty)$, $\delta > 0$, and $\kappa > \kappa' > 0$ we define the **bi-covering number profile** by

$$\text{BICOV}_{\alpha, \kappa, \kappa', \delta}(X, d_1, d_2, \mu) := \min_{\|\frac{d\mu'}{d\mu}\|_\infty \leq \alpha} \max_{\substack{U \subseteq X \\ \mu'(U) \geq \kappa}} \text{bicov}_{\kappa'}((U, d_1, d_2, \mu'), \delta).$$

This definition is quite involved, and clearly warrants some discussion.

An intuitive way to think about Definition 5.11 is in terms of a competition between two players, Max-er and Min-er. Given a compact pair-psm.p. space (X, d_1, d_2, μ) , Max-er and Min-er compete to produce a subset $U \subseteq X$. Max-er's goal is to maximize the resulting value of $\text{bicov}_{\kappa'}((U, d_1, d_2, \mu'), \delta)$ for some new auxiliary measure μ' , and Min-er's goal is to minimize it. They play as follows¹:

1. First, Min-er may choose any new measure $\mu' \in \text{Pr } X$, provided

$$\left\| \frac{d\mu'}{d\mu} \right\|_\infty \leq \alpha. \tag{15}$$

The natural choice to imagine here is $\mu' := \mu|_A$ for some $A \subseteq X$ with $\mu(A) \geq \alpha^{-1}$. We allow the relaxation to arbitrary measures satisfying (15) because it makes some later arguments smoother (and we work with $\|\cdot\|_\infty$, rather than any other norm, also as a matter of convenience).

¹Note that because the number of turns is limited to two, this is not a 'game' in the fully-fledged mathematical sense.

2. Second, Max-er chooses a subset $U \subseteq X$ for which $\mu'(U) \geq \kappa$. For instance, if $\mu' = \mu|_A$, then this is equivalent to $\mu(U \cap A) \geq \kappa\mu(A)$. So this choice by Max-er is constrained by Min-er's earlier choice of μ' : for instance, for any subset A of measure at least α^{-1} , Min-er is able to force Max-er to include a not-too-small piece of that subset in her choice of U .

(Implicitly, there is a third minimization turn implied by the definition of bicov , in which Min-er chooses a subset of U of measure at least κ' that can be covered most efficiently by bi-neighbourhoods. The flexibility of this last choice is also important in case Max-er's choice of U contains some unwieldy subset of measure less than $\kappa - \kappa'$, since Min-er is then not required to cover that portion of U .)

Let us motivate this idea by sketching how it repairs certain defects of its simpler relative in (13).

As suggested above, (13) can be used to give an isomorphism invariant of p.-p. systems. The problem seems to be that it is very difficult to compute, for two distinct reasons.

- Firstly, X could contain small subsets that have a heavy ‘pathological’ effect on the bi-covering numbers, in that they either decrease or increase them drastically. A drastic decrease is easy to visualize: imagine removing a tendrill of fairly small measure which is long and thin for both d_1 and d_2 . This possibility would already be dealt with by our requiring only a partial covering of X , up to a certain measure. However, as seen in Example 5.10, removing a subset can also *increase* bi-covering numbers, and I do not know how to rule out the possibility that removing a very small subset is responsible for a very large increase. We need a definition that is stable under this possibility as well.

Definition 5.11 overcomes this latter problem in the second turn of the competition above: it is in Max-er's interest to choose a subset that removes any ‘decreasing pathology’.

- Secondly, even if one is allowed to trim away pathologies of both the kinds above, the pair-psm.p. spaces

$$(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu)$$

can still be quite ‘inhomogeneous’: they can contain various large-measure subsets that exhibit a broad spectrum of different asymptotics for their bi-covering numbers. It could be difficult to work out how these different subsets contribute to an overall bi-covering number. This will be discussed further for the particular skew-products of Theorem A in Section 4.

To overcome this problem, Definition 5.11 allows Min-er a first turn in which he is allowed to restrict attention to any not-too-small subset – this should result in him cutting away the ‘bigger part’ of X from the point of view of bicov.

Crucially, the formulation of Definition 5.11 in terms of repeated optimization — that is, as a competition — gives a way to excise these problems that is intrinsic to the pair-psm.p.-space structure. This intrinsicity of BICOV will be key to its giving an isomorphism invariant of systems.

Understanding Definition 5.11 in terms of a competition will also help to guide us through the proofs of estimates on BICOV values later in the paper. To prove an upper bound, one imagines playing as Min-er with Max-er playing optimally, and to prove a lower bound, one imagines the reverse.

5.3 The new invariant

To define our new invariant in terms of BICOV, it is natural to focus on the metrics appearing in compact models. Theorem 3.4 gives such a model for any system, and we will soon show that two isomorphic compact models give the same invariant up to some natural equivalence. (However, it is sometimes convenient to use other pseudometrics on X for some comparison with the metric in a compact model, hence the decision to include general pseudometrics above.)

The marginal pair-m.p. spaces of different compact systems are related using the following extension of Lusin’s Theorem.

Lemma 5.12. *Let $\Phi : (X, d^X, \mu, T) \longrightarrow (Y, d^Y, \nu, S)$ be a Borel factor map of compact model p.-p. systems, and let $\varepsilon > 0$. Then there is an $L < \infty$ such that for all sufficiently large $N \in \mathbb{N}$ there is a compact subset $X_0 \subseteq X$ with $\mu(X_0) > 1 - \varepsilon$ and such that $\Phi|_{X_0}$ is continuous and (εN) -almost L -pair-Lipschitz from $(X_0, d_{[-N;0]}^X, d_{[0;N]}^X)$ to $(Y, d_{[-N;0]}^Y, d_{[0;N]}^Y)$.*

Proof. Let $\varepsilon_1 := \varepsilon / (2\text{diam}(Y, d^Y) + 1)$. By Lusin’s Theorem, there is a compact subset $X_1 \subseteq X$ such that $\mu(X_1) > 1 - \varepsilon_1^2$ and $\Phi|_{X_1}$ is continuous. That continuity implies that $\Phi|_{X_1}$ is also ε_1 -almost L -Lipschitz for some $L < \infty$.

Now let $N \in \mathbb{N}$, and for each $x \in X$ let

$$I_{1,x} := \{n \in [-N; 0) \mid T^n x \notin X_1\} \quad \text{and} \quad I_{2,x} := \{n \in [0; N) \mid T^n x \notin X_1\}.$$

Let

$$X_2 := \{x \in X \mid |I_{1,x} \cup I_{2,x}| \leq 2\varepsilon_1 N\}.$$

Since

$$\int_X |I_{1,x} \cup I_{2,x}| \mu(dx) = 2N\mu(X \setminus X_1) < 2\varepsilon_1^2 N,$$

Markov's Inequality implies $\mu(X \setminus X_2) < \varepsilon_1 \leq \varepsilon$.

Now suppose that $x, x' \in X_2$. Then

$$\begin{aligned} & d_{[-N;0]}^{\mathbf{Y}}(\Phi(x), \Phi(x')) \\ &= \sum_{n=-N}^{-1} d^Y(\Phi(T^n x), \Phi(T^n x')) \\ &\leq |I_{1,x} \cup I_{1,x'}| \text{diam}(Y, d^Y) + \sum_{n \in [-N;0] \setminus I_{1,x} \cup I_{1,x'}} d^Y(\Phi(T^n x), \Phi(T^n x')) \\ &\leq 2\varepsilon_1 N \text{diam}(Y, d^Y) + \varepsilon_1 N + L \sum_{n=-N}^{-1} d^X(T^n x, T^n x') \\ &= \varepsilon N + L d_{[-N;0]}^{\mathbf{X}}(x, x'), \end{aligned}$$

showing that $\Phi|_{X_2}$ is (εN) -almost L -Lipschitz. The analogous estimate holds also for $d_{[0;N]}^{\mathbf{X}}$ and $d_{[0;N]}^{\mathbf{Y}}$.

Finally, another appeal to Lusin's Theorem gives a further compact subset $X_0 \subseteq X_2$ such that $\Phi|_{X_0}$ is continuous, $\mu(X_0) > 1 - \varepsilon$, and the above almost Lipschitz bounds must still hold. \square

Proposition 5.13. *Suppose that $\Phi : (X, d^X, \mu, T) \longrightarrow (Y, d^Y, \nu, S)$ is a Borel factor map of compact model p.-p. systems, and that $\alpha > 1$, $\delta > 0$ and $\kappa > \kappa' > 0$. Then for every $\alpha_1 \in [1, \alpha)$ there are $\kappa_1 \in (\kappa', \kappa)$ and $\delta_1 \in (0, \delta)$ such that*

$$\text{BICOV}_{\alpha_1, \kappa_1, \kappa', \delta_1 N}(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu) \geq \text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(Y, d_{[-N;0]}^{\mathbf{Y}}, d_{[0;N]}^{\mathbf{Y}}, \nu)$$

for all sufficiently large N .

Proof. The parameters $\alpha_1 < \alpha$, $\kappa' < \kappa$ and δ are fixed. Choose ε so small that one has

$$\alpha\varepsilon < 1, \quad \alpha_1 < (1 - \alpha\varepsilon)\alpha, \quad \varepsilon < \delta/2, \quad \text{and} \quad \kappa_1 := (1 - \alpha\varepsilon)\kappa \in (\kappa', \kappa).$$

Let $L < \infty$ be given by Lemma 5.12 for this ε , and now choose δ_1 so small that $L\delta_1 + \varepsilon < \delta$.

Having chosen these parameters, and given any N which is sufficiently large for the conclusion of Lemma 5.12, set

$$m := \text{BICOV}_{\alpha_1, \kappa_1, \kappa', \delta_1 N}(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu).$$

We will imagine playing as Min-er in the competition described in Subsection 5.2 with the input space $(Y, d_{[-N;0]}^{\mathbf{Y}}, d_{[0;N]}^{\mathbf{Y}}, \nu)$.

By the definition of m , we may choose some $\mu' \in \text{Pr } X$ with $\|\text{d}\mu'/\text{d}\mu\|_\infty \leq \alpha_1$, and with the property that for every $U \subseteq X$ one has

$$\mu'(U) \geq \kappa_1 \implies \text{bicov}_{\kappa'}((U, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu'), \delta_1 N) \leq m. \quad (16)$$

Now, recalling our choice of ε , let $X_0 \subseteq X$ be given by Lemma 5.12, and then let $\mu'' := \mu'|_{X_0}$. Since $\mu'(X \setminus X_0) \leq \alpha\mu(X \setminus X_0) < \alpha\varepsilon$, we have

$$\left\| \frac{\text{d}\mu''}{\text{d}\mu} \right\|_\infty \leq \frac{\alpha_1}{1 - \alpha\varepsilon} < \alpha.$$

Also, if $U \subseteq X_0$ with $\mu''(U) \geq \kappa$, then

$$\mu'(U) \geq \mu'(X_0)\kappa > (1 - \alpha\varepsilon)\kappa = \kappa_1,$$

whereas for any $W \subseteq X_0$ one has $\mu''(W) = \mu'(W)/\mu'(X_0) \geq \mu'(W)$. Therefore (16) implies that also

$$\mu''(U) \geq \kappa \implies \text{bicov}_{\kappa'}((U, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu''), \delta_1 N) \leq m.$$

Finally, let $\nu'' := \Phi_*\mu''$, so this also satisfies $\|\text{d}\nu''/\text{d}\nu\|_\infty < \alpha$. This will be our choice of measure on Y . Given any $V \subseteq Y$ with $\nu''(V) \geq \kappa$, let $U := \Phi^{-1}(V) \cap X_0$. Then also $\mu''(U) \geq \kappa$, and Φ defines a c -almost L -pair-Lipschitz and measure-preserving map

$$(U, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu'') \longrightarrow (V, d_{[-N;0]}^{\mathbf{Y}}, d_{[0;N]}^{\mathbf{Y}}, \nu''),$$

so Lemma 5.9 gives

$$\begin{aligned} & \text{bicov}_{\kappa'}((U, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu''), \delta_1 N) \\ & \geq \text{bicov}_{\kappa'}((V, d_{[-N;0]}^{\mathbf{Y}}, d_{[0;N]}^{\mathbf{Y}}, \nu''), (L\delta_1 + c)N). \end{aligned}$$

Since $L\delta_1 + c \leq \delta$, this completes the proof. \square

Remark 5.14. *It is not clear how well bi-covering numbers behave under Cartesian products. However, one cannot hope for any nontrivial estimates for joinings. This can be seen from the result of [ST79] that any positive-entropy system is a joining of three Bernoulli factors. We will see later that Bernoulli systems gives trivial BICOV values, whereas some positive-entropy systems, such as nontrivial RWRs, do not — so the triviality of the former cannot give a bound on the latter. \triangleleft*

Definition 5.15. *The family of sequences*

$$\text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu), \quad N \geq 1,$$

parameterized by $\alpha \geq 1$, $\kappa > \kappa' > 0$ and $\delta > 0$, is the **bi-covering rate** of (X, d^X, μ, T) . Proposition 5.13 implies that it depends only on the isomorphism class of (X, μ, T) , up to the notion of equivalence implied by that proposition.

In the sequel, it will also be useful to compare the bi-covering rates of different pseudometrics defined on the same system. The following is immediate from the definition of bicov_\bullet .

Lemma 5.16. *Let (X, d^X, μ, T) be a compact model p.-p. system, let $M, \varepsilon > 0$, and let ρ be a totally bounded Borel pseudometric on X such that $d \leq M\rho + \varepsilon$. Then also*

$$\begin{aligned} \text{BICOV}_{\alpha, \kappa, \kappa', (M\delta + \varepsilon)N}(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu) \\ \leq \text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(X, \rho_{[-N;0]}^{\mathbf{X}}, \rho_{[0;N]}^{\mathbf{X}}, \mu) \end{aligned}$$

for all $\alpha > 1$, $\kappa > \kappa' > 0$, $\delta > 0$ and $N \in \mathbb{N}$. \square

Most often this will be used with $\rho(x, x') := 1_{\mathcal{P}(x) \neq \mathcal{P}(x')}$ for some finite measurable partition \mathcal{P} .

5.4 Two elementary examples

Before broaching the bi-covering rates of generalization RWRS systems, it will be instructive to analyze them in two rather simpler cases. This subsection is essentially a digression, and can be skipped without missing any of the proof of Theorem A.

5.4.1 Isometric systems

Proposition 5.17. *If T is an isometry of the compact metric space (X, d) and $\mu \in \text{Pr}^T X$ is ergodic, then*

$$\begin{aligned} \min_{\left\| \frac{d\mu'}{d\mu} \right\|_\infty \leq \alpha} \max_{\mu'(U) \geq \kappa} \text{cov}_{\kappa'}((U, d, \mu'), 2\delta) \\ \leq \text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu) \\ \leq \min_{\left\| \frac{d\mu'}{d\mu} \right\|_\infty \leq \alpha} \max_{\mu'(U) \geq \kappa} \text{cov}_{\kappa'}((U, d, \mu'), \delta) \end{aligned}$$

for all $\alpha \in (1, \infty)$, $\delta > 0$, $\kappa > \kappa' > 0$, and $N \in \mathbb{N}$.

Of course, the optimizations involved in these upper and lower bounds may still be non-trivial, and depend rather delicately on (X, d, μ) , but they do not involve N .

Proof. Since T is an isometry, one has $d \circ (T^{\times 2})^n = d$ for all n , and hence

$$d_{[-N;0]}^{\mathbf{X}} = d_{[0;N]}^{\mathbf{X}} = N \cdot d.$$

For any metric space (X, d) , any $\delta > 0$ and any $x \in X$, one has

$$B_\delta(x) \subseteq B_\delta(B_\delta(x)) \subseteq B_{2\delta}(x).$$

Therefore, for any $U \subseteq X$, one has

$$\begin{aligned} \text{cov}_{\kappa'}((U, d), 2\delta) &\leq \text{bicov}_{\kappa'}((U, d, d), \delta) \\ &= \text{bicov}_{\kappa'}((U, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}), \delta N) \leq \text{cov}_{\kappa'}((U, d), \delta). \end{aligned}$$

Now performing the optimization over μ' and U completes the proof. \square

5.4.2 Bernoulli systems

Proposition 5.18. *If (X, d^X, μ, T) is a compact model of a Bernoulli system, then*

$$\text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(X, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}}, \mu) = 1$$

for all sufficiently large N , for all $\alpha > 1$, $\kappa > \kappa' > 0$ and $\delta > 0$.

Thus, Propositions 5.17 and 5.18 show that both compact systems and Bernoulli systems have bi-covering rates that do not grow with N , even though they are in many ways ‘extreme opposites’ with regard to mixing behaviour.

The first step is an auxiliary result comparing marginal distributions over different time-intervals. Let \mathbf{X} be as above, let \mathcal{P} be any finite Borel partition of X , let $m := |\mathcal{P}|$ and let $\xi : X \rightarrow [m]$ be a finite-valued function generating \mathcal{P} . Given $\eta > 0$ and $N \in \mathbb{N}$, define

$$X_{N,\eta}^{\text{fat}} := \{x \in X \mid \xi_*^{[0;N]}(\mu|_{\mathcal{P}^{[-N;0]}(x)}) \ll_{\exp(\eta N), \eta} \xi_*^{[0;N]} \mu\}.$$

Lemma 5.19. *For every $\eta > 0$ one has*

$$\mu(X_{N,\eta}^{\text{fat}}) \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

Proof. Equation (2) and a special case of Lemma 3.3 give

$$\begin{aligned} \int_X D_{\text{KL}}(\xi_*^{[0;N]}(\mu_{|\mathcal{P}^{[-N;0]}(x)} | \xi_*^{[0;N]}\mu) \mu(dx) &= I_\mu(\mathcal{P}^{[0;N]}; \mathcal{P}^{[-N;0]}) \\ &= o(N) \quad \text{as } N \longrightarrow \infty. \end{aligned}$$

Therefore, by Markov's inequality, the sets

$$\{x \in X \mid D_{\text{KL}}(\xi_*^{[0;N]}(\mu_{|\mathcal{P}^{[-N;0]}(x)} | \xi_*^{[0;N]}\mu) \leq \eta^2 N - e^{-1}\}$$

have measure tending to 1 as $N \longrightarrow \infty$. By Lemma 2.5 with $C := \eta N$, these are contained in the sets $X_{N,\eta}^{\text{fat}}$. \square

The approximate absolute continuity in the definition of $X_{N,\eta}^{\text{fat}}$ will be used in conjunction with Theorem 4.3. Let X_N^{conc} be a sequence of high-probability Borel subsets of X such that $(X_N^{\text{conc}}, N^{-1}d_{[0;N]}^{\mathcal{P}}, \mu_{|X_N^{\text{conc}}})$ exhibits exponential measure concentration, as given by that theorem. Let $c(\delta) > 0$ be the exponential rate of concentration for this sequence for each radius $\delta > 0$. Also, if $I = [a, a+N] \subseteq \mathbb{Z}$, then let $X_I^{\text{conc}} := T^a(X_N^{\text{conc}})$. This is clearly \mathcal{P}^I -measurable.

Now given $\gamma, \delta > 0$, define

$$\begin{aligned} X_{N,\gamma,\delta}^{\text{loc.exp}} &:= \{x \in X_{[0;N]}^{\text{conc}} \mid \text{if } U \subseteq X \text{ and } \mu_{|\mathcal{P}^{[-N;0]}(x)}(U) \geq \gamma \text{ then} \\ &\quad \mu(B_{\delta N}^{d_{[0;N]}^{\mathcal{P}}}(U) \mid X_{[0;N]}^{\text{conc}}) > 1 - e^{-c(\delta)N}\}. \end{aligned}$$

Intuitively, $X_{N,\gamma,\delta}^{\text{loc.exp}}$ consists of those x such that if an event U is reasonably likely given the ‘past’ $\mathcal{P}^{[-N;0]}(x)$, then a small Hamming-neighbourhood around U for the ‘future’ $\mathcal{P}^{[0;N]}$ is very nearly the whole of $X_{[0;N]}^{\text{conc}}$.

Lemma 5.20. *For every $\gamma, \delta > 0$, one has*

$$\mu(X_{N,\gamma,\delta}^{\text{loc.exp}}) \longrightarrow 1 \quad \text{as } N \longrightarrow \infty.$$

Proof. Choose some $\eta < \min\{\gamma/2, c(\delta)\}$, which implies that

$$\frac{\gamma e^{-\eta N} (1 - 2\eta/\gamma)}{2} \geq e^{-c(\delta)N}$$

for all sufficiently large N .

Let

$$\begin{aligned} Y_N &:= \{x \in X_{[0;N]}^{\text{conc}} \mid \mu_{|\mathcal{P}^{[-N;0]}(x)}(X_{[0;N]}^{\text{conc}}) > 1 - \gamma/2 \\ &\quad \text{and } \xi_*^{[0;N]}(\mu_{|\mathcal{P}^{[-N;0]}(x)}) \ll_{\exp(\eta N), \eta} \xi_*^{[0;N]}\mu\}. \end{aligned}$$

Since

$$\mu(X_{[0;N]}^{\text{conc}}) = \int \mu|_{\mathcal{P}^{[-N;0]}(x)}(X_{[0;N]}^{\text{conc}}) \mu(dx) \longrightarrow 1 \quad \text{as } N \longrightarrow \infty,$$

Markov's Inequality, Lemma 5.19 and Theorem 4.3 imply that $\mu(Y_N) \longrightarrow 1$ as $N \longrightarrow \infty$.

We will show that $Y_N \subseteq X_{N,\gamma,\delta}^{\text{loc.exp}}$, so suppose that $x \in Y_N$ and $U \subseteq \mathcal{P}^{[-N;0]}(x)$ with $\mu|_{\mathcal{P}^{[-N;0]}(x)}(U) \geq \gamma$. Let $U' := U \cap X_{[0;N]}^{\text{conc}}$, so the definition of Y_N implies that $\mu|_{\mathcal{P}^{[-N;0]}(x)}(U') \geq \gamma/2$. Let $W := \mathcal{P}^{[0;N]}(U')$, so

$$B_{\delta N}^{d_{[0;N]}^{\mathcal{P}}}(W) = B_{\delta N}^{d_{[0;N]}^{\mathcal{P}}}(U') \subseteq B_{\delta N}^{d_{[0;N]}^{\mathcal{P}}}(U). \quad (17)$$

Now one has

$$\xi_*^{[0;N]}(\mu|_{U'}) \ll_{2/\gamma,0} \xi_*^{[0;N]}(\mu|_{\mathcal{P}^{[-N;0]}(x)}) \ll_{\exp(\eta N),\eta} \xi_*^{[0;N]}\mu,$$

and hence, by the rules in Lemma 2.2,

$$\xi_*^{[0;N]}(\mu|_{U'}) \ll_{2\exp(\eta N)/\gamma, 2\eta/\gamma} \xi_*^{[0;N]}\mu.$$

This implies that

$$\mu(U') \geq \frac{\gamma e^{-\eta N}}{2} (\mu|_{U'}(U') - 2\eta/\gamma) = \frac{\gamma e^{-\eta N} (1 - 2\eta/\gamma)}{2},$$

and this is at least $e^{-c(\delta)N}$ for N large enough, by our choice of η . Now (17) and the definition of $c(\delta)$ complete the proof. \square

Proof of Proposition 5.18. First fix $\delta > 0$, let \mathcal{P} be a finite Borel partition of X into sets of diameter less than $\delta/2$, and let $d^{\mathcal{P}}$ be the associated pseudometric: $d^{\mathcal{P}}(x, x') := 1_{\mathcal{P}(x) \neq \mathcal{P}(x')}$. It follows that

$$B_{\delta|I|}^{d_I^{\mathbf{X}}}(x) \supseteq B_{\delta|I|/2}^{d_I^{\mathcal{P}}}(x) \quad (18)$$

for any bounded discrete interval $I \subseteq \mathbb{Z}$ and any $x \in X$.

Now fix $\alpha \in (1, \infty)$ and $\kappa > \kappa' > 0$, and let $\gamma := \kappa/\alpha$ and $\varepsilon := (\kappa - \kappa')/\alpha$. We will show that for any sufficiently large N and any $U \subseteq X$ with $\mu(U) \geq \gamma$, there are many points $x \in U$ has the property that

$$\mu(U \cap B_{\delta N/2}^{d_{[0;N]}^{\mathcal{P}}}(U \cap \mathcal{P}^{[-N;0]}(x))) > \mu(U) - \varepsilon. \quad (19)$$

In view of (18), this implies the same lower bound for the (δN) -bi-neighbourhood of x in $(U, d_{[-N;0]}^{\mathbf{X}}, d_{[0;N]}^{\mathbf{X}})$, completing the proof.

To prove (19), observe that

$$\gamma \leq \mu(U) = \int_X \mu|_{\mathcal{P}^{[-N;0]}(x)}(U) \mu(dx) = \int_{X_{N,\gamma/2,\delta/2}^{\text{loc.exp}}} \mu|_{\mathcal{P}^{[-N;0]}(x)}(U) \mu(dx) + o(1)$$

as $N \rightarrow \infty$, so provided N is sufficiently large, there must be some $x \in X_{N,\gamma/2,\delta/2}^{\text{loc.exp}}$ such that

$$\mu|_{\mathcal{P}^{[-N;0]}(x)}(U) \geq \gamma/2.$$

Since this implies that $U \cap \mathcal{P}^{[-N;0]}(x) \neq \emptyset$, and the left-hand side of this last inequality depends only on the cell $\mathcal{P}^{[-N;0]}(x)$, we may move x within that cell if necessary to assume in addition that $x \in U$. However, by the definition of $X_{N,\gamma/2,\delta/2}^{\text{loc.exp}}$, these assumptions now imply

$$\begin{aligned} \mu(B_{\delta N/2}^{\mathcal{P}}(U \cap \mathcal{P}^{[-N;0]}(x))) \\ \geq \mu(B_{\delta N/2}^{\mathcal{P}}(U \cap \mathcal{P}^{[-N;0]}(x)) \mid X_{[0;N],\delta/2}^{\text{conc}}) - \mu(X \setminus X_{[0;N],\delta/2}^{\text{conc}}) \\ > 1 - e^{-c(\delta/2)N} - \mu(X \setminus X_{[0;N],\delta/2}^{\text{conc}}), \end{aligned}$$

and this is greater than $1 - \varepsilon$ for all sufficiently large N , implying (19). \square

5.5 Behaviour of the invariant for generalized RWRS systems

We will now formulate our main result for the bi-covering rate of generalized RWRS systems. It involves a certain universal function $\psi_{\text{BM}} : [1, \infty) \rightarrow (0, \infty]$, defined in terms of geometric features of Brownian sample paths.

Definition 5.21. *The function $\psi_{\text{BM}} : [1, \infty) \rightarrow (0, \infty]$ is defined as follows:*

$$\psi_{\text{BM}}(\alpha) := \inf \left\{ \psi \in (0, \infty] \mid W_{[0,1]}^{\otimes 2} \{ \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) \leq \psi \} \geq 1/\alpha \right\}.$$

It is easy to check that the random variable

$$\begin{aligned} C_0(0,1] \times C_0(0,1] &\longrightarrow [0, \infty) : (B, B') \mapsto \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) \\ &= \min \left\{ \sup_{0 \leq t \leq 1} B_t, \sup_{0 \leq t \leq 1} B'_t \right\} - \max \left\{ \inf_{0 \leq t \leq 1} B_t, \inf_{0 \leq t \leq 1} B'_t \right\} \end{aligned}$$

has an atomless distribution under $W^{\otimes 2}$. Using this and standard properties of Brownian motion, one easily verifies the following.

Lemma 5.22. *The function ψ_{BM} has the following properties:*

- $\psi_{\text{BM}}(1) = \infty$;
- $\psi_{\text{BM}}(\alpha) \in (0, \infty)$ for all $\alpha > 1$;
- ψ_{BM} is strictly decreasing;
- ψ_{BM} is continuous.

□

In terms of ψ_{BM} , our main result is as follows.

Theorem 5.23. *If (\mathbf{Y}, σ) is a well-distributed pair satisfying the Enhanced Invariance Principle, \mathbf{X} is an ergodic compact model flow, and $\alpha \in (1, \infty)$, then*

$$\sup_{\kappa > \kappa' > 0} \sup_{\delta > 0} \limsup_{N \rightarrow \infty} \frac{\log \text{BICOV}_{\alpha, \kappa, \kappa', \delta}(Y \times X, d_{[-N; 0]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}, d_{[0; N]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}, \nu \otimes \mu)}{\sqrt{N}} = \psi_{\text{BM}}(\alpha) h(\mathbf{X}).$$

(Proposition 5.13 already shows that this $\sup \limsup$ does not depend on the choice of compact metric models for \mathbf{Y} and \mathbf{X} as abstract p.-p. systems.) The reason for this somewhat delicate dependence on the properties of Brownian motion will become clear during the proof.

Remark 5.24. *I expect that the limit-supremum here is actually a limit, and that this requires only a slight enhancement of the proof of the lower bound given below. However, that enhancement seems to require rather heavier bookkeeping, so we do not pursue it in this paper.* ◁

Proof of Theorem A from Theorem 5.23. Suppose that $\mathbf{Y} \times_{\sigma} \mathbf{X}_1$ and $\mathbf{Y} \times_{\sigma} \mathbf{X}_2$ are two examples as in Theorem A, and that the former admits a factor map to the latter. Then we may take logarithms in the inequality of Proposition 5.13, divide by \sqrt{N} , and then deduce from Theorem 5.23 that $h(\mathbf{X}_1) \geq h(\mathbf{X}_2)$. □

The rest of this paper is given to the proof of Theorem 5.23. This will involve separate proofs of upper and lower bounds, the second being the more difficult direction.

6 The combinatorial basis of the upper bound

This section introduces a general tool which will underly the proof of the upper bound in Theorem 5.23. Although very elementary, it may be of interest in its own right. Subsection 6.2 also gives an easier outing for this tool, proving that the bi-covering rates of arbitrary p.-p. systems are always sublinear.

6.1 A bound using mutual information

For this subsection, fix a probability space (X, μ) . We will next develop ways to find an efficient covering of a ‘large’ (in terms of μ) portion of X using certain distinguished subsets, based on some other information about those subsets.

Our most basic result in this direction assumes that these special subsets are involved in a reasonably ‘smooth’ barycentric decomposition of μ .

Lemma 6.1. *Suppose that (X, μ) and (Z, ν) are standard Borel probability spaces, and that $z \mapsto \mu_z$ is a measurable family of finite Radon measures on X , uniformly bounded, such that*

$$\mu = \int_Z \mu_z \nu(dz) \quad \text{and} \quad \mu_z \ll_{M, \varepsilon} \mu \quad \forall z \in Z.$$

Suppose in addition that for each $z \in Z$, Y_z is a Borel subset of X for which $\mu_z(Y_z) = 1$. Then for every $\alpha < 1 - \varepsilon$ there is a subset $S \subseteq Z$ with

$$|S| \leq \frac{M}{1 - \alpha - \varepsilon} \quad \text{and} \quad \mu\left(\bigcup_{z \in S} Y_z\right) > \alpha.$$

Note that the measures μ_z are not required to be probability measures; this flexibility will be helpful shortly.

Proof. The set S is constructed by the following greedy recursion.

Suppose that $z_1, \dots, z_m \in Z$ have already been picked, where this is vacuous if $m = 0$. If $\mu\left(\bigcup_{i=1}^m Y_{z_i}\right) > \alpha$, then Stop and set $S := \{z_1, \dots, z_m\}$. Otherwise, let $U := X \setminus \bigcup_{i=1}^m Y_{z_i}$, and observe that

$$1 - \alpha \leq \mu(U) = \int_Z \mu_z(U) \nu(dz),$$

so there is some $z_{m+1} \in Z$ for which

$$\mu_{z_{m+1}}(U) = \mu_{z_{m+1}}(U \cap Y_{z_{m+1}}) \geq 1 - \alpha,$$

and hence

$$M\mu(U \cap Y_{z_{m+1}}) + \varepsilon \geq 1 - \alpha \implies \mu(U \cap Y_{z_{m+1}}) \geq \frac{1 - \alpha - \varepsilon}{M}.$$

This gives the choice of the next point z_{m+1} .

Having obtained z_1, \dots, z_m by the above algorithm, we have

$$\mu\left(\bigcup_{i=1}^m Y_{z_i}\right) = \sum_{i=1}^m \mu\left(Y_{z_i} \setminus \bigcup_{j=1}^{i-1} Y_{z_j}\right) \geq m \cdot \frac{1 - \alpha - \varepsilon}{M}.$$

This requires that $m \leq M/(1 - \alpha - \varepsilon)$, so the recursion must terminate in a set S containing at most this many points. The union of the corresponding supports must have measure greater than α , since this was the condition for termination. \square

Lemma 6.1 gives the covering conclusion that we will need later, but its assumption that $\mu_z \ll_{M,\varepsilon} \mu$ uniformly in z is stronger than we will meet directly. We will next turn it into an estimate closer to our applications. This begins with a useful way of ‘trimming’ a positive-measure subset U of a probability space (X, Σ, μ) relative to a finite measurable partition \mathcal{P} of X .

Definition 6.2. Let $\gamma \geq 0$, let (X, Σ, μ) be a probability space, and let $\mathcal{P} \subseteq \Sigma$ be a finite partition into positive-measure sets. A subset $V \in \Sigma$ is **locally γ -thick in \mathcal{P}** if for every $C \in \mathcal{P}$ one has

$$\text{either } C \cap V = \emptyset \quad \text{or} \quad \mu(V \mid C) \geq \gamma.$$

For clarity, note that \emptyset is locally γ -thick in every partition, for every γ .

The following lemma is an immediate consequence of Markov’s Inequality, but it will be worth having it ready to hand.

Lemma 6.3 (Trimming a set to a partition). *Let (X, Σ, μ) be a probability space, let $\mathcal{P} \subseteq \Sigma$ be a finite partition into positive-measure sets, let $U \in \Sigma$ with $\mu(U) > 0$, and let $\alpha \in (1/2, 1)$. Then the subset*

$$V := \bigcup_{\substack{C \in \mathcal{P} \\ \mu(U \mid C) \geq (1 - \alpha)\mu(U)}} (U \cap C) \subseteq U$$

satisfies $\mu(V) \geq \alpha\mu(U)$ and is locally $((1 - \alpha)\mu(U))$ -thick in \mathcal{P} . \square

The above definition and lemma have an obvious generalization to local thickness relative to a Borel map $\pi : X \rightarrow Y$ and a given disintegration of μ over π , but this will not be needed.

Now assume that \mathcal{S} and \mathcal{T} are two fixed finite Borel partitions of (X, μ) .

Proposition 6.4. *Let $I := I_\mu(\mathcal{S}; \mathcal{T})$ and suppose that $\alpha \in (0, 1]$ and $\eta \in (0, \alpha)$. Then for every Borel $U \subseteq X$ with $\mu(U) \geq \alpha$, there is a subset $S \subseteq U$ with*

$$\log |S| \lesssim_{\alpha, \eta} I + 1 \quad \text{and} \quad \mu(U \cap \mathcal{T}(U \cap \mathcal{S}(S))) > \mu(U) - \eta$$

(where the notation in the first inequality indicates that the bound depends on α and η but not otherwise on the choice of U).

The connection between the mutual-information bound assumed here and the hypothesis of Lemma 6.1 will result from Lemma 2.5.

Proof. Let $\psi : X \longrightarrow A$ be a map to a finite set that generates the partition \mathcal{T} . Then equation (2) gives

$$I = \int_X D_{\text{KL}}(\psi_*(\mu_{|\mathcal{T}(x)}) \mid \psi_*\mu) \mu(dx).$$

Step 1. Choose $\zeta := \eta/3$, and observe that this implies

$$(1 - \zeta)^2(\mu(U) - \zeta) > (1 - 2\zeta)(\mu(U) - \zeta) > \mu(U) - 3\zeta = \mu(U) - \eta.$$

Now let $D := I/\zeta$. Applying Markov's Inequality to the integral above gives that the set

$$X_1 := \{x \mid D_{\text{KL}}(\psi_*(\mu_{|\mathcal{T}(x)}) \mid \psi_*\mu) \leq D\}$$

has

$$\mu(X_1) \geq 1 - I/D = 1 - \zeta.$$

Letting $U_1 := U \cap X_1$, it follows that $\mu(U_1) \geq \mu(U) - \zeta$.

On the other hand, for any $C \in (0, \infty)$, Lemma 2.5 gives

$$X_1 \subseteq \{x \mid \psi_*(\mu_{|\mathcal{T}(x)}) \ll_{e^C, (D+e^{-1})/C} \psi_*\mu\}. \quad (20)$$

Step 2. Now let $\gamma := \zeta\mu(U_1) \geq \zeta(\mu(U) - \zeta)$, and apply Lemma 6.3 to find some $V \subseteq U_1$ with

$$\mu(V) \geq (1 - \zeta)\mu(U_1) \geq (1 - \zeta)(\mu(U) - \zeta) > (\mu(U) - \eta)/(1 - \zeta)$$

and which is locally γ -thick in \mathcal{T} .

Step 3. The decomposition of μ into the measures $\mu_{|\mathcal{T}(x)}$ may be conditioned on V and pushed forward under ψ to obtain

$$\psi_*(\mu_{|V}) = \frac{1}{\mu(V)}\psi_*(1_V \cdot \mu) = \frac{1}{\mu(V)} \int_X \psi_*(1_V \cdot \mu_{|\mathcal{T}(x)}) \mu(dx) \quad (21)$$

(being aware that the measures inside the right-hand integral may now not be probability measures). Let $B := \psi(V) \subseteq A$, so the above pushforward measures are all supported on B . Applying (20), it follows that any $x \in V \subseteq X_1$ satisfies

$$\psi_*(1_V \cdot \mu_{|\mathcal{T}(x)}) \leq 1_B \cdot \psi_*(\mu_{|\mathcal{T}(x)}) \ll_{e^C, (D+e^{-1})/C} 1_B \cdot \psi_*\mu.$$

On the other hand, since V is locally γ -thick in \mathcal{T} , for any $b \in B$ one has $\mu(V \cap \psi^{-1}\{b\}) \geq \gamma \mu(\psi^{-1}\{b\})$. Therefore

$$1_B \cdot \psi_* \mu \leq \frac{1}{\gamma} \psi_*(1_V \cdot \mu) = \frac{\mu(V)}{\gamma} \psi_*(\mu|_V).$$

Combining this with the preceding inequalities, and recalling that $\mu(V) \geq \mu(U) - \eta \geq \alpha - \eta$, we obtain

$$\frac{1}{\mu(V)} \psi_*(1_V \cdot \mu|_{\mathcal{S}(x)}) \ll_{e^C/\gamma, (D+e^{-1})/C(\alpha-\eta)} \psi_*(\mu|_V).$$

Step 4. This relates the integral and integrands in (21), and so puts us in position to apply Lemma 6.1. The family of measures is $(1/\mu(V))\psi_*(1_V \cdot \mu|_{\mathcal{S}(x)}) \in \text{Pr } A$ for $x \in X$, and for each x the relevant supporting subset is $\psi(V \cap \mathcal{S}(x)) \subseteq A$. To carry out this application, it remains to choose the constant C . Let $\varepsilon := \zeta/2$, and now let

$$C := \frac{D + e^{-1}}{\varepsilon(\alpha - \eta)} = \frac{2(I/\zeta + e^{-1})}{\zeta(\alpha - \eta)} \lesssim_{\alpha, \eta} I + 1.$$

For the assumptions of Lemma 6.1 we must set $M := e^C$, and can then obtain some $S \subseteq V$ such that $|S| \leq M/(\zeta - \varepsilon) = 2M/\zeta$ and

$$\begin{aligned} \mu|_V(\mathcal{T}(V \cap \mathcal{S}(S))) &= \psi_*(\mu|_V)(\psi(V \cap \mathcal{S}(S))) > 1 - \zeta \\ \implies \mu(U \cap \mathcal{T}(U \cap \mathcal{S}(S))) &\geq \mu(V \cap \mathcal{T}(V \cap \mathcal{S}(S))) > (1 - \zeta)\mu(V) \geq \mu(U) - \eta. \end{aligned}$$

This gives a subset S contained in X , but not necessarily in U . However, one can discard any $x \in S$ such that $U \cap \mathcal{S}(x) = \emptyset$ without disrupting these estimates, and any remaining x can be replaced by an element of $U \cap \mathcal{S}(x)$ to give an element of U with the same \mathcal{S} -cell. We may therefore take $S \subseteq U$, as required. \square

One further generalization of the above result will be important later. To formulate it, we now posit a third partition \mathcal{R} such that $\mathcal{R} \leq \mathcal{S} \wedge \mathcal{T}$.

Proposition 6.5. *Let $I := I_\mu(\mathcal{S}; \mathcal{T} | \mathcal{R})$ and suppose that $\alpha \in (0, 1]$ and $\eta \in (0, \alpha)$. Then for every Borel $U \subseteq X$ with $\mu(U) \geq \alpha$ there is a subset $S \subseteq U$ such that*

$$\log \frac{|S|}{|\mathcal{R}|} \lesssim_{\alpha, \eta} I + 1 \quad \text{and} \quad \mu(U \cap \mathcal{T}(U \cap \mathcal{S}(S))) > \mu(U) - \eta.$$

Proof. By discarding the union of all μ -negligible cells of \mathcal{R} , we may assume that all cells have positive measure.

Step 1. Again let $\zeta := \eta/3$, and let $J := I/\zeta$. Recall that

$$I = \int_X \mathbf{I}_{\mu|_{\mathcal{R}(x)}}(\mathcal{S}; \mathcal{T}) \mu(dx),$$

and let

$$\mathcal{R}_0 := \{C \in \mathcal{R} \mid \mathbf{I}_{\mu|_C}(\mathcal{S}; \mathcal{T}) \leq J\},$$

so Markov's Inequality gives

$$\mu\left(\bigcup \mathcal{R}_0\right) \geq 1 - I/J = 1 - \zeta.$$

Letting $U_0 := U \cap \bigcup \mathcal{R}_0$, it follows that $\mu(U_0) \geq \mu(U) - \zeta$.

Step 2. Now let $\gamma := \zeta\mu(U_0)$, and apply Lemma 6.3 to obtain $V \subseteq U_0$ with $\mu(V) \geq (1 - \zeta)\mu(U_0)$ and which is locally γ -thick in \mathcal{R} . Since $V \subseteq U_0$, we know that $V \cap \mathcal{R} = V \cap \mathcal{R}_0$. Let $\mathcal{R}_1 := \{C \in \mathcal{R}_0 \mid C \cap V \neq \emptyset\}$.

Step 3. Finally, for each $C \in \mathcal{R}_1$, consider the probability space $(C, \mu|_C)$ and the two partitions $\mathcal{S} \cap C$ and $\mathcal{T} \cap C$. Since $\mathcal{R}_1 \subseteq \mathcal{R}_0$, we know that

$$\mathbf{I}_{\mu|_C}(\mathcal{S} \cap C; \mathcal{T} \cap C) \leq J,$$

while Step 2 guarantees that $\mu|_C(V) \geq \gamma \geq \zeta(\mu(U) - \zeta) \geq \zeta(\alpha - \zeta)$. This last lower bound depends only on α and η , as does ζ , so we may apply Proposition 6.4 within each of these conditioned probability spaces to obtain subsets $S_C \subseteq V \cap C$ such that

$$\log |S_C| \lesssim_{\alpha, \eta} J + 1 \lesssim_{\eta} I + 1 \quad \forall C \in \mathcal{R}_1,$$

while

$$\mu|_C(V \cap C \cap \mathcal{T}(V \cap C \cap \mathcal{S}(S_C))) > \mu|_C(V) - \zeta.$$

Let $S := \bigcup_{C \in \mathcal{R}_1} S_C$. Then

$$\log \frac{|S|}{|\mathcal{R}|} \leq \max_{C \in \mathcal{R}_1} \log |S_C| \lesssim_{\alpha, \eta} I + 1,$$

and

$$\begin{aligned}
\mu(U \cap \mathcal{T}(U \cap \mathcal{S}(S))) &\geq \mu(V \cap \mathcal{T}(V \cap \mathcal{S}(S))) \\
&= \sum_{C \in \mathcal{R}_1} \mu(C) \mu|_C(V \cap \mathcal{T}(V \cap \mathcal{S}(S))) \\
&\geq \sum_{C \in \mathcal{R}_1} \mu(C) \mu|_C(V \cap C \cap \mathcal{T}(V \cap C \cap \mathcal{S}(S_C))) \\
&> \sum_{C \in \mathcal{R}_1} \mu(C) (\mu|_C(V) - \zeta) \\
&\geq \mu(V) - \zeta \geq (1 - \zeta)(\mu(U) - \zeta) - \zeta \geq \mu(U) - \eta.
\end{aligned}$$

□

6.2 An upper bound for general systems

This is another digressive subsection, but it also offers a warm-up to the upper bound in Theorem 5.23. It will use Proposition 6.4 to prove a bi-covering-rate upper bound for arbitrary p.-p. systems. It proves our first concrete relation between bi-covering rates and mutual information.

Proposition 6.6. *Let (X, d^X, μ, T) be a compact model p.-p. system. For any $\alpha \geq 1$, $\kappa > \kappa' > 0$, and $\delta > 0$, there is a finite Borel partition \mathcal{P} of X such that*

$$\log \text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(X, d_{[-N;0]}^X, d_{[0;N]}^X, \mu) \lesssim_{\kappa, \kappa'} I_\mu(\mathcal{P}^{[-N;0]}, \mathcal{P}^{[0;N]}) + 1$$

as $N \rightarrow \infty$.

Proof. It suffices to treat the case $\alpha = 1$, in which BICOV admits no choice of new measures on X . For this case, let \mathcal{P} be any finite Borel partition of X into cells of diameter less than δ . Then

$$B_{\delta|I|}^{d^X}(x) \supseteq \mathcal{P}^I(x) \quad \forall x \in X \text{ and finite } I \subseteq \mathbb{Z},$$

and so also

$$B_{\delta N}^{d_{[0;N]}^X}(U \cap B_{\delta N}^{d_{[-N;0]}^X}(S)) \supseteq \mathcal{P}^{[0;N]}(U \cap \mathcal{P}^{[-N;0]}(S)) \quad \forall S \subseteq U.$$

Now Proposition 6.4 promises some $S \subseteq U$ such that

$$\mu(U \cap \mathcal{P}^{[0;N]}(U \cap \mathcal{P}^{[-N;0]}(S))) \geq \kappa'$$

and also

$$\log |S| \lesssim_{\kappa, \kappa'} I_\mu(\mathcal{P}^{[-N;0]}, \mathcal{P}^{[0;N]}) + 1.$$

□

Combined with Lemma 3.3, this immediately gives the following.

Proposition 6.7. *For any compact model p - p . system (X, d^X, μ, T) one has*

$$\log \text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(X, d_{[-N; 0]}^{\mathbf{X}}, d_{[0; N]}^{\mathbf{X}}, \mu) = o_{\kappa, \kappa', \delta}(N)$$

as $N \rightarrow \infty$. □

Proposition 6.7 begs the following question.

Question 6.8. *Can the upper bound in Proposition 6.7 be improved to any fixed sub-linear function?*

This seems highly unlikely, but it could be interesting to see examples of (X, d^X, μ, T) whose log-bi-covering rates come arbitrarily close to linear. Conjecture 10.1 will propose some systems that could achieve rate $N^{1-\varepsilon}$ for any $\varepsilon > 0$.

7 The upper bound

This section proves the upper bound in Theorem 5.23. The proof is based on the covering estimates of the previous section, similarly to the proof of Proposition 6.7. The key is to replace certain balls for the metrics $d_I^{\mathbf{Y} \times_{\sigma} \mathbf{X}}$, $I \subseteq \mathbb{Z}$, with the cells of associated partitions, and then prove a mutual information bound for an application of Proposition 6.5. Most of the delicacy here will be in choosing the partitions that approximate the metrics.

In principle, one feels that proofs of these results should be possible directly in terms of the metrics $d_I^{\mathbf{Y} \times_{\sigma} \mathbf{X}}$, without this switch to partitions. However, I suspect that would require much thornier estimates in several places.

7.1 Estimating balls in the skew-product metric

Let \mathcal{P} be the time-zero partition of the SFT Y , as previously. Also, let us normalize the metric on X to assume that $\text{diam}(X, d^X) \leq 1$.

Lemma 7.1. *For every $\varepsilon > 0$ there is a $p \in \mathbb{N}$ such that for all $N \in \mathbb{N}$ and $y, y' \in Y$ one has the following implication:*

$$\mathcal{P}^{[-p; N+p]}(y) = \mathcal{P}^{[-p; N+p]}(y') \implies \max_{n \in [0; N]} |\sigma_n^y - \sigma_n^{y'}| < \varepsilon.$$

Proof. Since $\sigma : Y \rightarrow \mathbb{R}$ is Hölder continuous, there are some $b < \infty$ and $\beta \in (0, 1)$ such that

$$\begin{aligned} \mathcal{P}^{[-p, N+p)}(y) &= \mathcal{P}^{[-p, N+p)}(y') \\ \implies |\sigma(S^i y) - \sigma(S^i y')| &< b\beta^{\min\{i+p, N+p-i\}} \quad \forall i \in [0; N]. \end{aligned}$$

Summing over i , this gives

$$|\sigma_n^y - \sigma_n^{y'}| \leq \sum_{i=0}^{n-1} |\sigma(S^i y) - \sigma(S^i y')| \leq 2 \sum_{i \geq p} b\beta^i \leq \frac{2b\beta^p}{1-\beta} \quad \forall n \in [0; N],$$

which is less than ε provided p is large enough. \square

The following re-write of the above lemma will be useful in Subsection 9.3.

Corollary 7.2. *For every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $N \in \mathbb{N}$ and $y, y' \in Y$ one has the following implication:*

$$\max_{n \in [0; N)} d^Y(S^n y, S^n y') < \delta \implies \max_{n \in [0; N)} |\sigma_n^y - \sigma_n^{y'}| < \varepsilon.$$

\square

Lemma 7.3. *For any $\delta > 0$ there are $p \in \mathbb{N}$ and a finite Borel partition \mathcal{Q} of X such that the following holds. If $I \in \text{Int}(\mathbb{R})$, if $y \in Y$ satisfies*

$$\sigma_{[0; N)}^y + [-1, 1] \subseteq I,$$

and if $x \in X$, then

$$B_{\delta N}^{d_{[0; N)}^{Y \times \sigma^X}}(y, x) \supseteq \mathcal{P}^{[-p, N+p)}(y) \times \mathcal{Q}^{I \cap \mathbb{Z}}(x),$$

and analogously with $[-N; 0)$ in place of $[0; N)$ throughout.

Proof. Since $T : \mathbb{R} \curvearrowright (X, d^X)$ is continuous and $\delta > 0$, any partition \mathcal{Q} of X into cells of sufficiently small diameter has the property that

$$\forall x, x' \in X, \quad \mathcal{Q}(x) = \mathcal{Q}(x') \implies d^X(T^t x, T^t x') < \delta/3 \quad \forall t \in [-1, 1]. \quad (22)$$

Fix a finite Borel partition \mathcal{Q} with this property.

Next, using again the continuity of T , choose $\varepsilon > 0$ so small that

$$\sup_{|t| \leq \varepsilon} \sup_{x \in X} d^X(x, T^t x) < \delta/3. \quad (23)$$

Lastly, choose p as given by Lemma 7.1 for this value of ε . Increase p further if necessary so that also $2^{-p+2} < \delta/3$.

Now let $y \in Y$, let $I \in \text{Int}(\mathbb{R})$ with $\sigma_{[0;N]}^y + [-1, 1] \subseteq I$, and let $x \in X$. Suppose that

$$(y', x') \in \mathcal{P}^{[-p; N+p]}(y) \times \mathcal{Q}^{I \cap \mathbb{Z}}(x).$$

The definition of d^Y gives

$$d^Y(S^n y, S^n y') \leq 2^{-p+2} < \delta/3 \quad \forall n \in [0; N],$$

and the choice of p gives

$$|\sigma_n^y - \sigma_n^{y'}| \leq \varepsilon \quad \forall n \in [0; N].$$

Therefore, by (23),

$$\begin{aligned} d_{[0;N]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}((y, x), (y', x')) &= d_{[0;N]}^{\mathbf{Y}}(y, y') + \sum_{n=0}^{N-1} d^X(T^{\sigma_n^y} x, T^{\sigma_n^{y'}} x') \\ &\leq \delta N/3 + \sum_{n=0}^{N-1} (d^X(T^{\sigma_n^y} x, T^{\sigma_n^y} x') + d^X(T^{\sigma_n^y} x', T^{\sigma_n^{y'}} x')) \\ &\leq \left(2\delta/3 + \max_{n \in [0;N]} d^X(T^{\sigma_n^y} x, T^{\sigma_n^y} x')\right) N. \end{aligned}$$

Finally, since $\mathcal{Q}^{I \cap \mathbb{Z}}(x) = \mathcal{Q}^{I \cap \mathbb{Z}}(x')$, the property (22) gives that

$$d^X(T^t x, T^t x') < \delta/3 \quad \forall t \in (I \cap \mathbb{Z}) + [-1, 1].$$

In particular, this holds for $t = \sigma_n^y$ for $n \in [0; N]$. □

7.2 Completion of the upper bound

Now fix arbitrary $\delta > 0$, $\alpha \in (1, \infty)$ and $\kappa > \kappa' > 0$. Let $\psi := \psi_{\text{BM}}(\alpha)$. For the upper bound in Theorem 5.23, it will suffice to show that

$$\begin{aligned} \text{BICOV}_{\alpha, \kappa, \kappa', \delta N}(Y \times X, d_{[-N;0]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}, d_{[0;N]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}, \nu \otimes \mu) \\ \leq \exp((h(\mathbf{X}) + \varepsilon)(\psi + \varepsilon)\sqrt{N} + \varepsilon\sqrt{N}) \end{aligned}$$

as $N \rightarrow \infty$, for every $\varepsilon > 0$.

Having chosen δ , let $p \in \mathbb{N}$ and let \mathcal{Q} be a finite Borel partition of X as given by Lemma 7.3. These will now also be fixed for the rest of the section. Concerning the system (X, μ, T^1) and its partition \mathcal{Q} , if $\varepsilon > 0$ and $I \in \text{Int}(\mathbb{R})$ is a bounded

interval, then let $X_{I,\varepsilon}^{\text{SM}}$ denote the set of ‘typical’ points for the partition $\mathcal{Q}^{I \cap \mathbb{Z}}$ according to the Shannon-McMillan Theorem 3.1.

The next step will be to introduce certain auxiliary subsets and partitions of X and $Y \times X$.

Lemma 7.4. *For each $\varepsilon > 0$ there are finitely many pairs of intervals*

$$(I_1, J_1), \dots, (I_k, J_k) \in \text{Int}(\mathbb{R}) \times \text{Int}(\mathbb{R})$$

such that

$$0 < \mathcal{L}^1(I_i \cap J_i) < \psi + \varepsilon \quad \forall i = 1, 2, \dots, k$$

and for which the following holds. For each $N \in \mathbb{N}$ there are pairwise-disjoint Borel subsets $W_N^1, \dots, W_N^k \subseteq Y$ such that

- i) $\nu(W_N^1 \cup \dots \cup W_N^k) > 1/\alpha$ for all sufficiently large N , and
- ii) for every $y \in W_N^i$ one has

$$\sigma_{[-N;0]}^y + [-1, 1] \subseteq \sqrt{N}I_i \quad \text{and} \quad \sigma_{[0;N]}^y + [-1, 1] \subseteq \sqrt{N}J_i.$$

Proof. Consider the set

$$W^0 := \{(B, B') \in C_0(0, 1] \times C_0(0, 1] \mid 0 < \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) < \psi + \varepsilon/3\}.$$

Definition 5.21 and Lemma 5.22 give that

$$W_{[0,1]}^{\otimes 2}(W_0) = W_{[0,1]}^{\otimes 2} \{ \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) < \psi + \varepsilon/3 \} > 1/\alpha.$$

Since $W_{[0,1]}^{\otimes 2}$ is inner-regular with respect to compact sets, it follows that for some $k \in \mathbb{N}$ one can find

- pairs of intervals $(I_i, J_i) \in \text{Int}(\mathbb{R})^2$ for $i = 1, 2, \dots, k$ such that

$$0 < \mathcal{L}^1(I_i \cap J_i) < \psi + \varepsilon$$

- and pairwise-disjoint Borel subsets $W^1, \dots, W^k \subseteq W^0$

such that

- i) $W^1 \cup \dots \cup W^k$ is open,
- ii) $W_{[0,1]}^{\otimes 2}(W^1 \cup \dots \cup W^k) > 1/\alpha$, and

iii) for every $(B, B') \in W^i$ one has

$$B_{[0,1]} + [-\varepsilon/3, \varepsilon/3] \subseteq I_i \quad \text{and} \quad B'_{[0,1]} + [-\varepsilon/3, \varepsilon/3] \subseteq J_i.$$

Let

$$W_N^i := \{y \in Y \mid (\text{traj}_{-N}(\sigma^y), \text{traj}_N(\sigma^y)) \in W^i\} \quad \text{for each } i = 1, \dots, k.$$

Since $W^1 \cup \dots \cup W^k$ is open in $C_0(0, 1] \times C_0(0, 1]$, the Portmanteau Theorem and Theorem 3.10 imply that $\nu(W_N^1 \cup \dots \cup W_N^k) > 1/\alpha$ for all sufficiently large N . The desired conclusion (ii) then holds provided also $1/\sqrt{N} < \varepsilon/3$. \square

Keeping the notation of the preceding lemma, now let $Z_N^i := W_N^i \times X_{\sqrt{N}(I_i \cap J_i), \varepsilon}^{\text{SM}}$ for each $i = 1, 2, \dots, k$, and let

$$Z_N := Z_N^1 \cup \dots \cup Z_N^k, \quad Z_N^c := (Y \times X) \setminus Z_N, \quad \text{and} \quad \mathcal{Z}_N := \{Z_N^1, \dots, Z_N^k, Z_N^c\}.$$

For these sets Z_N we have

$$(\nu \otimes \mu)(Z_N) \geq \nu(W_N) - \max_{i \leq k} \mu(X \setminus X_{\sqrt{N}(I_i \cap J_i) \cap \mathbb{Z}}^{\text{SM}}),$$

so this is still greater than $1/\alpha$ for all sufficiently large N , by Theorem 3.1 and the fact that $|\sqrt{N}(I_i \cap J_i) \cap \mathbb{Z}| \rightarrow \infty$ for each i .

We next introduce the further partitions of $Y \times X$ that will enable an approximation of our bi-neighbourhoods.

First, for each $i = 1, 2, \dots, k$, define

$$\begin{aligned} \mathcal{R}_N^i &:= \{\emptyset, Y\} \otimes \mathcal{Q}^{\sqrt{N}(I_i \cap J_i) \cap \mathbb{Z}}, \\ \mathcal{S}_N^i &:= \mathcal{P}^{[-N-p; p]} \otimes \mathcal{Q}^{\sqrt{N}I_i \cap \mathbb{Z}}, \quad \text{and} \\ \mathcal{T}_N^i &= \mathcal{P}^{[-p; N+p]} \otimes \mathcal{Q}^{\sqrt{N}J_i \cap \mathbb{Z}}. \end{aligned}$$

These different partitions can be adapted to the cells Z_N^i as follows: let \mathcal{R}_N , \mathcal{S}_N and \mathcal{T}_N be the partitions that refine \mathcal{Z}_N , all contain Z_N^c as a single cell, and satisfy

$$\mathcal{R}_N \cap Z_N^i = \mathcal{R}_N^i \cap Z_N^i, \quad \mathcal{S}_N \cap Z_N^i = \mathcal{S}_N^i \cap Z_N^i \quad \text{and} \quad \mathcal{T}_N \cap Z_N^i = \mathcal{T}_N^i \cap Z_N^i$$

for all $i = 1, 2, \dots, k$ and $N \in \mathbb{N}$. Clearly

$$\mathcal{Z}_N \leq \mathcal{R}_N \leq \mathcal{S}_N \wedge \mathcal{T}_N.$$

Now Lemma 7.3 and the properties of the sets W_N^i given by Lemma 7.4 imply that

$$\mathcal{S}_N(y, x) \subseteq B_{\delta N}^{\mathbf{Y} \times_{[-N; 0]} \mathbf{X}}(y, x) \quad \text{and} \quad \mathcal{T}_N(y, x) \subseteq B_{\delta N}^{\mathbf{Y} \times_{[0; N]} \mathbf{X}}(y, x) \quad \forall (y, x) \in Z_N. \quad (24)$$

Lemma 7.5. *With \mathcal{R}_N as above, one has*

$$|\mathcal{R}_N| \leq \exp \left((h(\mathbf{X}) + \varepsilon)(\psi + \varepsilon)\sqrt{N} + o(\sqrt{N}) \right) \quad \text{as } N \longrightarrow \infty.$$

Proof. The definition above gives

$$|\mathcal{R}_N| = 1 + \sum_{i=1}^k |\mathcal{R}_N \cap Z_N^i| = 1 + \sum_{i=1}^K |\mathcal{Q}^{\sqrt{N}(I_i \cap J_i) \cap \mathbb{Z}} \cap X_{\sqrt{N}(I_i \cap J_i), \varepsilon}^{\text{SM}}|.$$

As $N \longrightarrow \infty$, the number of summands on the right-hand side here is fixed, and their cardinalities are bounded by

$$(h(\mathbf{X}) + \varepsilon)|\sqrt{N}(I_i \cap J_i) \cap \mathbb{Z}| \leq (h(\mathbf{X}) + \varepsilon)(\psi + \varepsilon)\sqrt{N}$$

by the definition of $X_{\sqrt{N}(I_i \cap J_i), \varepsilon}^{\text{SM}}$. □

Lemma 7.6. *With \mathcal{R}_N , \mathcal{S}_N and \mathcal{T}_N as above, one has*

$$I_{\nu \otimes \mu}(\mathcal{S}_N; \mathcal{T}_N | \mathcal{R}_N) = o(\sqrt{N}) \quad \text{as } N \longrightarrow \infty.$$

Proof. The definition of conditional mutual information gives

$$\begin{aligned} I_{\nu \otimes \mu}(\mathcal{S}_N; \mathcal{T}_N | \mathcal{R}_N) &= (\nu \otimes \mu)(Z_N^c) I_{(\nu \otimes \mu)|_{Z_N^c}}(\mathcal{S}_N; \mathcal{T}_N | \mathcal{R}_N) \\ &\quad + \sum_{i=1}^k (\nu \otimes \mu)(Z_N^i) I_{(\nu \otimes \mu)|_{Z_N^i}}(\mathcal{S}_N; \mathcal{T}_N | \mathcal{R}_N). \end{aligned}$$

Now the definition of \mathcal{R}_N , \mathcal{S}_N and \mathcal{T}_N gives that the first term here is zero, and the remaining sum is equal to

$$\sum_{i=1}^k (\nu \otimes \mu)(Z_N^i) I_{(\nu \otimes \mu)|_{Z_N^i}}(\mathcal{S}_N^i; \mathcal{T}_N^i | \mathcal{R}_N^i).$$

Applying Lemma 2.4, this is bounded by

$$\sum_{i=1}^k (\log 2 + I_{\nu \otimes \mu}(\mathcal{S}_N^i; \mathcal{T}_N^i | \mathcal{R}_N^i)),$$

and within each of these summands, an application of Lemma 2.3, Corollary 3.9 and Lemma 3.3 gives

$$\begin{aligned} &I_{\nu \otimes \mu}(\mathcal{S}_N^i; \mathcal{T}_N^i | \mathcal{R}_N^i) \\ &= I_{\nu}(\mathcal{P}^{[-N-p;p]}; \mathcal{P}^{[-p;N+p]}) + I_{\mu}(\mathcal{Q}^{\sqrt{N}I_i \cap \mathbb{Z}}; \mathcal{Q}^{\sqrt{N}J_i \cap \mathbb{Z}} | \mathcal{Q}^{\sqrt{N}(I_i \cap J_i) \cap \mathbb{Z}}) \\ &= O(1) + o(|(\sqrt{N}J_i \cap \mathbb{Z}) \setminus (\sqrt{N}I_i \cap \mathbb{Z})|) = o(\sqrt{N}) \end{aligned}$$

as $N \longrightarrow \infty$. □

Proof of upper bound in Theorem 5.23. Letting $\lambda_N := (\nu \otimes \mu)|_{Z_N}$, the lower bound on $(\nu \otimes \mu)(Z_N)$ implies that $\|\mathrm{d}\lambda_N/\mathrm{d}(\mu \otimes \nu)\|_\infty < \alpha$ for all sufficiently large N . This will be Min-er's choice of new measures on $Y \times X$: that is, we will prove that

$$\begin{aligned} \max_{\substack{U \subseteq Y \times X \\ \lambda_N(U) \geq \kappa}} \text{bicov}_{\kappa'}((U, d_{[-N;0]}^{\mathbf{Y} \times \sigma \mathbf{X}}, d_{[0;N]}^{\mathbf{Y} \times \sigma \mathbf{X}}, \lambda_N), \delta N) \\ \leq \exp((\psi + \varepsilon)(h(\mathbf{X}) + \varepsilon)\sqrt{N} + o(\sqrt{N})) \quad (25) \end{aligned}$$

as $N \rightarrow \infty$. In estimating this maximum we may assume that $U \subseteq Z_N$, for any remainder $U \setminus Z_N$ carries none of the measure λ_N and so does not need to be covered. After assuming that $U \subseteq Z_N$, we have the containments in (24) for all $(y, x) \in U$.

Now Proposition 6.5 gives a subset $S \subseteq U$ such that

$$\log \frac{|S|}{|\mathcal{R}_N|} \lesssim_{\alpha, \kappa, \kappa''} \mathbf{I}_{\nu \otimes \mu}(\mathcal{S}_N; \mathcal{T}_N | \mathcal{R}_N) + 1 \quad (26)$$

and

$$\begin{aligned} (\nu \otimes \mu)(U \cap \mathcal{T}_N(U \cap \mathcal{S}_N(S))) &> (\nu \otimes \mu)(U) - \kappa''/\alpha \\ \implies \lambda_N(U \setminus \mathcal{T}_N(U \cap \mathcal{S}_N(S))) &< \kappa'' \\ \implies \lambda_N(U \cap \mathcal{T}_N(U \cap \mathcal{S}_N(S))) &> \kappa - \kappa'' = \kappa'. \end{aligned}$$

Applying Lemmas 7.5 and 7.6 to the estimate (26), it follows that

$$|S| \leq \exp((h(\mathbf{X}) + \varepsilon)(\psi + \varepsilon)\sqrt{N} + o(\sqrt{N}))$$

as $N \rightarrow \infty$. This completes the proof. \square

The proof of the matching lower bound in Theorem 5.23 will require more delicate analysis than the above, and will occupy the remainder of this paper.

Remark 7.7. *The previous lemma is the point at which we make crucial use of the very fast mixing of \mathcal{P} via Corollary 3.9. In fact, that conclusion is slightly stronger than we need: it would suffice in the above argument to know that*

$$\mathbf{I}_\nu(\mathcal{P}^{[-N;0]}, \mathcal{P}^{[0;N]}) = o(\sqrt{N}).$$

However, if one had instead

$$\mathbf{I}_\nu(\mathcal{P}^{[-N;0]}, \mathcal{P}^{[0;N]}) = \Omega(\sqrt{N}),$$

then this would disrupt all of the subsequent estimates, since it would turn out that the mutual information of $\mathcal{P}^{[-N;0)}$ and $\mathcal{P}^{[0;N)}$ is of the same order as the entropy contained in that part of the scenery that has been visited over both $[-N;0)$ and $[0;N)$.

It would be interesting to know whether the upper bound proved above holds if one knows only that \mathbf{Y} is a Bernoulli system, that \mathcal{P} is a generating partition, and that σ satisfies the Invariance Principle. Of course, in this case one may choose an independent generating partition \mathcal{P}' of \mathbf{Y} , but now \mathcal{P}' may not give precise enough control over the cocycle σ , in the sense of Lemma 7.1, to give a usable analog of Lemma 7.3. This is why the argument above needs the precise relation between \mathcal{P} and σ that holds in a well-distributed pair. \triangleleft

8 Meandering of cocycles and discrete Cantor sets

For the lower bound in Theorem 5.23, we will imagine playing as Max-er in the competition described in Subsection 5.2. The key to Max-er's strategy will be an 'inverse theorem', asserting that for most pairs of sceneries and walk trajectories, if the pair of resulting strings produced by the RWRS process are close, then it must be because of some 'structural' similarity involving the sceneries alone.

This section introduces the structures that appear in this notion of similarity. The first of these is a class of special subsets of certain intervals $[0;N)$ on which our cocycle-trajectories are often (approximately) injective. These will be introduced after a discussion of another pre-requisite property of cocycle-trajectories.

Throughout this section (\mathbf{Y}, σ) will be a well-distributed pair. Recall that this means $Y \subseteq A^{\mathbb{Z}}$ is a mixing SFT with a Hölder-potential Gibbs measure ν , and $\sigma : Y \rightarrow \mathbb{R}$ is a one-sided Hölder non-coboundary with $\int \sigma d\nu = 0$ and effective variance 1. We assume also that (\mathbf{Y}, σ) satisfies the Enhanced Invariance Principle (Definition 3.13) with some mollifier φ . Finally, fix $\ell \geq \|\sigma\|_{\infty}$ large enough that $\text{spt } \varphi \subseteq [-\ell, \ell]$.

Many of the arguments of this section are adapted from similar steps in [Kal82], or their re-telling in [dHS97].

8.1 Two useful estimates

The starting point for this section is a pair of basic estimates on the distribution of our cocycle-trajectories.

Lemma 8.1. *In the above setting, one has*

$$\nu\{\sigma_N^y \in I\} \lesssim_{\mathbf{Y}, \sigma} \frac{\max\{\mathcal{L}^1(I), 1\}}{\sqrt{N}} \quad \forall I \in \text{Int}(\mathbb{R}), N \in \mathbb{N}.$$

In particular,

$$\nu\{|\sigma_N^y| \leq a\sqrt{N}\} \lesssim_{\mathbf{Y},\sigma} \max\left\{a, \frac{1}{\sqrt{N}}\right\} \quad \forall a \in (0, \infty), N \in \mathbb{N}.$$

Proof. Let $I = [a, b]$. Theorem 3.12 gives

$$\begin{aligned} \nu\{a \leq \sigma_N^y \leq b\} &= \nu\{\sigma_N^y \leq b\} - \nu\{\sigma_N^y \leq a\} \\ &\lesssim_{\mathbf{Y},\sigma} \mathrm{N}(-\infty, b/\sqrt{N}) - \mathrm{N}(-\infty, a/\sqrt{N}) + \frac{1}{\sqrt{N}}, \end{aligned}$$

so the result now follows from the smoothness of the Gaussian density. The second conclusion follows by taking $I = [-a, a]$. \square

Lemma 8.2. *In the above setting,*

$$\nu\left\{\max_{n \in [0;N)} |\sigma_n^y| \geq b\sqrt{N}\right\} \lesssim_{\mathbf{Y},\sigma} \frac{1}{b^2} \quad \forall b \in (0, \infty), N \in \mathbb{N}.$$

The proof of Lemma 8.2 is a little more involved. It begins with the following property of Gibbs measures.

Lemma 8.3 (Four-fold exponential mixing). *With (\mathbf{Y}, σ) as above, there are some $c < \infty$ and $\gamma \in (0, 1)$ such that*

$$\left| \int \sigma \cdot (\sigma \circ S^p) \cdot (\sigma \circ S^{p+q}) \cdot (\sigma \circ S^{p+q+r}) \, d\nu \right| \leq c\gamma^{\max\{p,r\}} \quad \forall p, q, r \in \mathbb{N} \cup \{0\}.$$

Proof. The definition of a Gibbs measure via equation (7) is invariant under time-reversal, as is the conclusion of the present lemma. It therefore suffices to find c and γ such that

$$\left| \int \sigma \cdot (\sigma \circ S^p) \cdot (\sigma \circ S^{p+q}) \cdot (\sigma \circ S^{p+q+r}) \, d\nu \right| \leq c\gamma^r \quad \forall p, q, r \in \mathbb{N}.$$

In this proof, let us re-scale σ so that $\|\sigma\|_\infty \leq 1$, and let L_ψ be a normalized Ruelle-Perron-Frobenius operator whose adjoint has invariant measure ν^- , as in Subsection 3.3.

Now let $k := \lfloor r/3 \rfloor$. Since σ is Hölder, there $b_1 < \infty$ and $\beta_1 \in (0, 1)$ such that

$$|\sigma(y) - \sigma(y')| \leq b_1 \beta_1^{d(y,y')},$$

and so

$$\|\sigma - \mathbb{E}_\nu(\sigma \mid \mathcal{P}^{[-k;k]})\|_\infty \leq b_1 \beta_1^k \leq (b_1/\beta_1)(\sqrt[3]{\beta_1})^r.$$

Letting $\sigma' := E_\nu(\sigma \mid \mathcal{P}^{[-k;k]})$, it follows that

$$\begin{aligned} & \left| \int \sigma \cdot (\sigma \circ S^p) \cdot (\sigma \circ S^{p+q}) \cdot (\sigma \circ S^{p+q+r}) d\nu \right| \\ & \leq \left| \int \sigma' \cdot (\sigma' \circ S^p) \cdot (\sigma' \circ S^{p+q}) \cdot (\sigma' \circ S^{p+q+r}) d\nu \right| + (b_1/\beta_1)(\sqrt[3]{\beta_1})^r. \end{aligned}$$

It therefore suffices to give an exponentially-decaying bound on the first term here. Since the whole integral is S -invariant, we may reduce instead to the case of σ' being $\mathcal{P}^{[-2k,0]}$ -measurable (hence, in particular, one-sided). Recalling that

$$E_\nu(\sigma' \circ S^r \mid \mathcal{P}^{(-\infty;0]}) = L_\psi^r \sigma'$$

for one-sided functions σ' , this leads to

$$\int \sigma' \cdot (\sigma' \circ S^p) \cdot (\sigma' \circ S^{p+q}) \cdot (\sigma' \circ S^{p+q+r}) d\nu = \int \sigma' \cdot (\sigma' \circ S^p) \cdot (\sigma' \circ S^{p+q}) \cdot (L_\psi^r \sigma' \circ S^{p+q}) d\nu.$$

Finally, because L_ψ has spectral radius less than 1 on any space of mean-zero, one-sided Hölder functions (see again [PP90, Theorem 2.2]), there are $b_2 < \infty$ and $\beta_2 \in (0, 1)$ such that

$$\left\| L_\psi^r \sigma' - \int \sigma' d\nu \right\|_\infty \leq b_2 \beta_2^k,$$

so this completes the proof. \square

Corollary 8.4 (Fourth-moment bound). *With (\mathbf{Y}, σ) as above, one has*

$$\|\sigma_N^y\|_{L^4(\nu)} \lesssim_{\mathbf{Y}, \sigma} \sqrt{N} \quad \forall N \in \mathbb{N}.$$

Proof. Let c and γ be as given by the preceding lemma. Expanding the power inside the L^4 -norm gives

$$\begin{aligned} \|\sigma_N^y\|_{L^4(\nu)}^4 &= \sum_{n_1, n_2, n_3, n_4 \in [0; N]} \int \sigma(S^{n_1}y) \sigma(S^{n_2}y) \sigma(S^{n_3}y) \sigma(S^{n_4}y) \nu(dy) \\ &\lesssim \sum_{p, q, r \geq 0, p+q+r < N} \int \sigma(y) \sigma(S^p y) \sigma(S^{p+q} y) \sigma(S^{p+q+r} y) \nu(dy) \\ &\leq c \sum_{p, q, r \geq 0, p+q+r < N} \gamma^{\max\{p, r\}} \\ &\lesssim_\gamma c N^2. \end{aligned}$$

\square

Proof of Lemma 8.2. Whenever $m, n \in [0; N)$ with $m \leq n$, the previous corollary gives

$$\int |\sigma_n^y - \sigma_m^y|^4 \nu(dy) = \int |\sigma_{n-m}^{S^m y}|^4 \nu(dy) \lesssim_{Y, \sigma} |n - m|^2.$$

This moment bound is strong enough to enable a standard chaining argument for controlling $\max_{n \in [0; N)} |\sigma_n^y|$. A suitable quantitative version is given by Billingsley as [Bil68, Theorem 12.2]. The bound above is the hypothesis of that theorem with (in his notation) parameters $\gamma = 4$, $\alpha = 2$ and $u_i = 1$ for all i , and its conclusion becomes

$$\nu\left\{\max_{n \in [0; N)} |\sigma_n^y| \geq b\sqrt{N}\right\} \lesssim_{Y, \sigma} \frac{N^2}{(b\sqrt{N})^4} = \frac{1}{b^4},$$

which is actually stronger than we require. \square

Remark 8.5. *The proof of [Bil68, Theorem 12.2] is really a quantitative implementation of Kolmogorov's classical proof that Brownian motion has a continuous version (see, for instance, [Kal02, Theorem 3.23]). The full sequence of arguments above — from a mixing result (Lemma 8.3), to a fourth-moment bound (Corollary 8.4), to an application of Kolmogorov's method — are essentially the steps taken by Bunimovich and Sinai in [BS81, Section 4] for their proof that the laws of the random variables $\text{traj}_N(\sigma^y)$ on (Y, ν) form a tight sequence in $\text{Pr } C[0, 1]$.*

An easy extension of Lemma 8.3 and Corollary 8.4 gives

$$\|\sigma_N^y\|_{L^{2p}(\nu)} \lesssim_{Y, \sigma, p} \sqrt{N} \quad \forall p \in \mathbb{N}.$$

The fourth moment is the simply smallest with which the method of Kolmogorov can be applied. \triangleleft

8.2 Meandering of cocycle-trajectories

Definition 8.6. *Let $\ell, \alpha > 0$, and let $M, L \in \mathbb{N}$. Let $a \in \mathbb{Z}$, let $I := [a; a + LM)$, and let*

$$\mathcal{C} = \{[a + iM; a + (i + 1)M) \mid i \in \{0, \dots, L - 1\}\}$$

be the partition of this interval into length- M subintervals.

For $y \in Y$, the cocycle-trajectory σ^y is (α, ℓ) -meandering over (I, \mathcal{C}) if the following holds:

$$\forall \mathcal{J} \subseteq \mathcal{C} \text{ with } |\mathcal{J}| \geq \alpha L \quad \exists J, J' \in \mathcal{J} \text{ such that } \text{dist}(\sigma_J^y, \sigma_{J'}^y) > 2\ell.$$

This property gives a sense in which the cocycle-trajectory σ^y has many well-separated images of intervals from \mathcal{C} . Importantly, once M is large, well-distributed pairs have a strong lower bound on the probability of this occurring.

Proposition 8.7. *If (Y, σ) is well-distributed, then there is a $C < \infty$ such that for all $\ell \in (0, \infty)$ and all sufficiently large $M \in \mathbb{N}$, the following holds for all $L \in \mathbb{N}$ and all $\alpha > 0$:*

$$\nu\{\sigma^y \text{ is } (\alpha, \ell)\text{-meandering over } ([0; LM), \mathcal{C})\} \geq 1 - \frac{C}{L^{1/3}\alpha^2},$$

where \mathcal{C} is the partition of $[0; LM)$ into subintervals of length M .

Remarks 8.8. (1.) *This is essentially the lower bound denoted by \spadesuit in [dHS97], except that they require only $\sigma_j^y \cap \sigma_{j'}^y = \emptyset$, rather than separation by 2ℓ . Both the proof and the later applications of this proposition roughly follow their paper, except that we work throughout with well-distributed pairs.*

In fact, [dHS97] allows a more general lower bound of the form $1 - \frac{C}{L^\gamma \alpha^2}$ for some $\gamma > 0$. The specific value $\gamma = 1/3$ arises for random walks in the domain of attraction of a Gaussian distribution, but a smaller value may be needed for a heavier-tailed walk, such as one in the domain of attraction of a p -stable law for some $p \in [1, 2)$. Their main results also apply to such random walks, hence their need for this generality, but our focus on well-distributed pairs precludes it. We will return to this point in Subsection 10.2.

(2.) *It will be very important that the value of M at which this inequality starts to hold does not depend on L or α .* \triangleleft

Proof of Proposition 8.7. This proof is essentially as in paragraphs (2.5) and (2.6) of [Kal82].

First, $L^{1/3}\alpha^2 \leq (\alpha L)^2$ for all $\alpha > 0$ and $L \in \mathbb{N}$. Therefore, provided C is sufficiently large, the desired bound is vacuous for small values of αL , and we may henceforth assume $\alpha L \geq 2$, and hence $\binom{\lceil \alpha L \rceil}{2} \geq (\alpha L)^2/4$.

Fix $i, j \in [L]$ with $j > i$, and suppose $\beta \in (0, 1/2)$ (it will be optimized later). Then Lemmas 8.1 and 8.2 give

$$\begin{aligned} & \nu\{\text{dist}(\sigma_{[iM; (i+1)M)}^y, \sigma_{[jM; (j+1)M)}^y) \leq (j-i)^\beta \sqrt{M}\} \\ & \leq \nu\{|\sigma_{(j-i)M}^y| \leq 2(j-i)^\beta \sqrt{M}\} + 2\nu\left\{\max_{n \in [0; M)} |\sigma_n^y| \geq (j-i)^\beta \sqrt{M}\right\} \\ & \lesssim_{Y, \sigma} \max\left\{\frac{(j-i)^\beta}{\sqrt{(j-i)}}, \frac{1}{\sqrt{(j-i)M}}\right\} + \frac{1}{(j-i)^{2\beta}} = \frac{(j-i)^\beta}{\sqrt{(j-i)}} + \frac{1}{(j-i)^{2\beta}}. \end{aligned}$$

Choosing $\beta := 1/6$, this last bound becomes $1/(j-i)^{1/3}$. Therefore, provided $\sqrt{M} > 2\ell$,

$$\begin{aligned}
& \nu\{\sigma^y \text{ is not } (\alpha, \ell)\text{-meandering over } ([0; LM), \mathcal{C})\} \\
& \leq \nu\{\exists J \subseteq [L] \text{ with } |J| \geq \alpha L \text{ such that } \forall i, j \in J \text{ distinct one has} \\
& \quad \text{dist}(\sigma_{[iM, (i+1)M)}^y, \sigma_{[jM, (j+1)M)}^y) \leq |j-i|^{1/6} \sqrt{M}\} \\
& \leq \nu\left\{\exists \text{ at least } \binom{\lceil \alpha L \rceil}{2} \text{ pairs } i, j \in [L] \text{ such that } j > i \text{ and} \right. \\
& \quad \left. \text{dist}(\sigma_{[iM, (i+1)M)}^y, \sigma_{[jM, (j+1)M)}^y) \leq |j-i|^{1/6} \sqrt{M}\right\} \\
& \leq \frac{1}{\binom{\lceil \alpha L \rceil}{2}} \sum_{1 \leq i < j \leq L} \nu\{\text{dist}(\sigma_{[iM, (i+1)M)}^y, \sigma_{[jM, (j+1)M)}^y) \leq (j-i)^{1/6} \sqrt{M}\} \\
& \lesssim_{\mathbf{Y}, \sigma} \frac{1}{\alpha^2 L^2} \sum_{1 \leq i < j \leq L} \frac{1}{(j-i)^{1/3}} \lesssim \frac{L \cdot L^{2/3}}{\alpha^2 L^2} = \frac{1}{\alpha^2 L^{1/3}},
\end{aligned}$$

as required. \square

8.3 Regularity for occupation measures

We next introduce the consequence that we will need of the Enhanced Invariance Principle (Definition 3.13). Our application of this principle will be essentially the same as made by Aaronson in [Aar12], for the special case of random walks: see the ‘Local Time Lemma’ and Lemma 4 of that paper.

Lemma 8.9 (Smoothness of typical occupation measures). *Suppose that (\mathbf{Y}, σ) satisfies the Enhanced Invariance Principle. Let φ and ℓ be as at the beginning of this section. For every $\varepsilon > 0$ there is an $M < \infty$ such that*

$$\nu\{\varphi \star \gamma_{[0;N)}^y \sim_{M,\varepsilon} \mathbb{U}_{B_\ell(\sigma_{[0;N)}^y)}\} \geq 1 - \varepsilon$$

for all sufficiently large N .

Proof. Because L^B is a.s. a continuous function with support *equal* to the positive-length compact interval $B_{[0,1]}$ ([Kal02, Corollary 22.18]), for any $\varepsilon > 0$ there are $M < \infty$ and finitely many $K_1, \dots, K_r \in \text{Int}(\mathbb{R})$ such that $\mathcal{L}^1(K_s) > 1/(M\varepsilon)$ for each s , and such that the set

$$\begin{aligned}
U &:= \bigcup_{s=1}^r \left\{ (K, f) \in \text{Int}(\mathbb{R}) \times C_c(\mathbb{R}) \mid \right. \\
& \quad K_s \subseteq \text{interior}(K) \subseteq K \subseteq K_s + (-1/2M, 1/2M), \\
& \quad \left. 2/M < f|_{K_s} < M/2, \text{ and } \int_{K_s} f > 1 - \varepsilon \right\}
\end{aligned}$$

satisfies

$$W_{[0,1]} \{B \mid (B_{[0,1]}, L^B) \in U\} > 1 - \varepsilon.$$

This last subset of $C_0(0, 1]$ is easily checked to be open. On the other hand, the Enhanced Invariance Principle implies that

$$(B_\ell(\sigma_{[0;N]}^y)/\sqrt{N}, \varphi \star \gamma_{[0;N]}^y(\sqrt{N}(\cdot))) \xrightarrow{\text{law}} (B_{[0,1]}, L^B)$$

as $(\text{Int}(\mathbb{R}) \times C_c(\mathbb{R}))$ -valued random variables. Therefore the Portmanteau Theorem implies that

$$\nu \{ (B_\ell(\sigma_{[0;N]}^y)/\sqrt{N}, \varphi \star \gamma_{[0;N]}^y(\sqrt{N}(\cdot))) \in U \} > 1 - \varepsilon$$

for all sufficiently large N . Since $\varphi \star \gamma_{[0;N]}^y(\sqrt{N}(\cdot))$ is always non-negative and has integral equal to 1, this membership of U implies that

$$\varphi \star \gamma_{[0;N]}^y \sim_{M,\varepsilon} \mathbb{U}_{B_\ell(\sigma_{[0;N]}^y)},$$

as required. \square

Based on the preceding results, we can introduce certain sets that occur with high probability as the relevant time scale tends to ∞ . With φ and ℓ as before, and for $[a; b] \subseteq \mathbb{Z}$, set

$$Y_{[a;b]}^{\text{spread}} := \{y \mid [\sigma_a^y - (b-a)^{1/3}, \sigma_a^y + (b-a)^{1/3}] \subseteq B_\ell(\sigma_{[a;b]}^y)\}.$$

Given also $\varepsilon > 0$ and $M < \infty$, set

$$Y_{[a;b],M,\varepsilon}^{\text{smooth}} := \{y \mid \varphi \star \gamma_{[a;b]}^y \sim_{M,\varepsilon} \mathbb{U}_{B_\ell(\sigma_{[a;b]}^y)}\}.$$

The results above show that

$$\nu(Y_{[a;b]}^{\text{spread}}) \longrightarrow 1 \quad \text{as } b-a \longrightarrow \infty \quad (27)$$

(indeed, this would follow if the exponent $1/3$ were replaced with any value less than $1/2$), and that

$$\nu(Y_{[a;b],M,\varepsilon}^{\text{smooth}}) \longrightarrow 1 \quad (28)$$

as $b-a \longrightarrow \infty$ and then $M \longrightarrow \infty$.

8.4 A discrete filtration

We will next define recursively some sequences of auxiliary parameters and sets.

Firstly, let

$$L_d := \lceil (d+1)^{18} \rceil \quad \forall d \geq 1$$

(the reason for the exponent 18 will emerge shortly). Let $N_0 := 1$, and then let

$$N_{d+1} := L_{d+1} N_d \quad \forall d \geq 0.$$

A crude application of Stirling's Inequality gives

$$\log N_d \sim (d+1) \log(d+1) \quad \forall d \geq 1. \quad (29)$$

In addition, let

$$\alpha_d := \frac{1}{(d+1)^2}$$

and

$$\kappa_{r,d} := \prod_{i=r+1}^d (1 - \alpha_i) \quad \text{whenever } 1 \leq r \leq d.$$

The series $\sum_i \alpha_i$ converges, and therefore $\kappa_{r,d}$ tends to a limit $\kappa_{r,\infty} \in [0, 1)$ as $d \rightarrow \infty$, and then $\kappa_{r,\infty} \uparrow 1$ as $r \rightarrow \infty$.

Also, for each $d \geq 0$, let \mathcal{D}_d be the partition of \mathbb{Z} into the discrete intervals

$$Q_i^d := [iN_d; (i+1)N_d) \quad \text{for } i \in \mathbb{Z}.$$

Given $t \in \mathbb{Z}$, let $Q^d(t)$ denote the cell of \mathcal{D}_d that contains t , so $Q^d(t) = Q_i^d$ if and only if $iN_d \leq t < (i+1)N_d$.

Letting ℓ be as before, for each $d \geq 1$ and $t \in \mathbb{Z}$ let

$$Y_{d,t}^{\text{mndr}} := \{y \in Y \mid \sigma^y \text{ is } (\alpha_d, \ell)\text{-meandering over } (Q^d(t), \mathcal{D}_{d-1} \cap Q^d(t))\}$$

(this is essentially a repeat of the definition of the set ' θ'_d ' in [dHS97, Section 6.1]; interestingly, our argument does not seem to need any analog of the sets ' θ'''_d ' from that paper). Observe that $Y_{d,t}^{\text{mndr}}$ actually depends on t only through $Q^d(t)$, hence only on $i := \lfloor t/N_d \rfloor$. For this i , one has

$$Y_{d,t}^{\text{mndr}} = S^{-iN_d}(Y_{d,0}^{\text{mndr}}),$$

and so Proposition 8.7 gives

$$\begin{aligned} \nu(Y_{d,t}^{\text{mndr}}) &= \nu(Y_{d,0}^{\text{mndr}}) > 1 - \frac{C}{L_d^{1/3} \alpha_d^2} \\ &\geq 1 - \frac{C}{(d+1)^{18/3} \frac{1}{(d+1)^4}} = 1 - \frac{C}{(d+1)^2}, \end{aligned} \quad (30)$$

with C as in that proposition, for all d and t . (This explains the choice of the exponent 18: we will soon need the error term at the end here to be summable in d .)

Given $1 \leq s \leq d$ and $Q \in \mathcal{D}_d$, let

$$H_{s,Q}^{\text{mndr}}(y) := \{t \in Q \mid y \in Y_{s,t}^{\text{mndr}}\}.$$

For a given y , this is the set of times $t \in Q$ such that σ^y ‘behaves well’, in the sense of Proposition 8.7, over the interval $Q^s(t)$. Since $Y_{s,t}^{\text{mndr}}$ depends on t only through $Q^s(t)$, the set $H_{s,Q}^{\text{mndr}}(y)$ is a union of cells from $\mathcal{D}_s \cap Q$: it is equal to

$$\bigcup \{[iN_s; (i+1)N_s) \mid i \in \mathbb{Z}, [iN_s; (i+1)N_s) \subseteq Q, \text{ and } S^{iN_s}y \in Y_{s,0}^{\text{mndr}}\}.$$

The analysis below will need cocycle-trajectories that are simultaneously ‘well-behaved’ over most intervals $Q^s(t) \subseteq Q \in \mathcal{D}_d$ on all sufficiently large scales up to some d . A high probability of such cocycle-trajectories is given by the following lemma.

Lemma 8.10. *If $1 \leq r < d$ and $Q \in \mathcal{D}_d$, then*

$$\int_Y \left| \bigcap_{s=r+1}^d H_{s,Q}^{\text{mndr}}(y) \right| \nu(dy) \geq \left(1 - C \sum_{s=r+1}^d \frac{1}{(s+1)^2}\right) N_d$$

with C as in Proposition 8.7.

Proof. For each fixed $s \in [1; d]$, the bound (30) gives

$$\begin{aligned} \int_Y |Q \setminus H_{s,Q}^{\text{mndr}}(y)| \nu(dy) &= \sum_{t \in Q} \nu\{y \mid t \notin H_{s,Q}^{\text{mndr}}(y)\} \\ &= N_d \nu(Y \setminus Y_{s,0}^{\text{mndr}}) < \frac{CN_d}{(s+1)^2}, \end{aligned}$$

and therefore

$$\begin{aligned} \int_Y \left| \bigcap_{s=r+1}^d H_{s,Q}^{\text{mndr}}(y) \right| \nu(dy) &\geq N_d - \sum_{s=r+1}^d \int_Y |Q \setminus H_{s,Q}^{\text{mndr}}(y)| \nu(dy) \\ &> \left(1 - C \sum_{s=r+1}^d \frac{1}{(s+1)^2}\right) N_d. \end{aligned}$$

□

We can introduce other sets of times in a fixed large interval Q at which a cocycle-trajectory is behaving well in one sense or another. Fix again $1 \leq r < d$ and $Q \in \mathcal{D}_d$. Then we set

$$H_{r,Q}^{\text{spread}}(y) := \{t \in Q \mid y \in Y_{Q^r(t)}^{\text{spread}}\}$$

and

$$H_{r,Q,M,\varepsilon}^{\text{smooth}}(y) := \{t \in Q \mid y \in Y_{Q^r(t),M,\varepsilon}^{\text{smooth}}\}.$$

Now we may combine the estimates (27) and (28) with Lemma 8.10 and Markov's Inequality to conclude the following.

Corollary 8.11. *For every $\eta > 0$ and $\varepsilon > 0$ there are $r, M \in \mathbb{N}$ such that, for every $d > r$ and every $Q \in \mathcal{D}_d$, one has*

$$\nu\left\{y \in Y \mid \left|H_{r,Q}^{\text{spread}}(y) \cap H_{r,Q,M,\varepsilon}^{\text{smooth}}(y) \cap \bigcap_{s=r+1}^d H_{s,Q}^{\text{mndr}}(y)\right| \geq (1-\eta)N_d\right\} > 1-\eta.$$

□

The largeness of the intersection of times appearing here implies that the cocycle-trajectory σ^y enjoys several different useful properties simultaneously on most of Q . The conjunction of all of these properties will play a rôle in our later analysis of generalized RWRS systems.

8.5 Discrete Cantor sets and approximate injectivity

This subsection will focus on the intersection $\bigcap_{s=r+1}^d H_{s,Q}^{\text{good}}(y)$, as appears in Corollary 8.11. Provided it is large enough, one can find special, highly-structured subsets inside it on which σ^y is approximately injective.

Definition 8.12 (Discrete Cantor sets). *Let $d \in \mathbb{N}$. A **discrete Cantor set of depth d** is an indexed family $(t_\omega)_{\omega \in \{0,1\}^d}$ of points in \mathbb{R} , and it is **proper** if they are all distinct.*

*In addition, given $D_1 \geq D_2 \geq \dots \geq D_d > 0$, these are **gap upper bounds** for the discrete Cantor set if*

$$|t_\omega - t_{\omega'}| \leq D_{i+1} \quad \text{whenever } i < d \text{ and } \omega_j = \omega'_j \ \forall j \leq i.$$

Given $K \in \text{Int}(\mathbb{R})$, $d \in \mathbb{N}$ and $D = (D_1 \geq \dots \geq D_d)$ as above, we will let $\text{DCS}_{d,D}(K) \subseteq K^{\{0,1\}^d}$ denote the collection of all discrete Cantor sets of depth d , contained in K , and having gap upper bounds given by D .

Our first result about discrete Cantor sets is an estimate on their ‘number’; or, more correctly, their covering number for some natural metric. We will endow $\text{DCS}_{d,D}(K)$ with the metric d_{DCS} obtained from the norm $\|\cdot\|_\infty$ on $\mathbb{R}^{\{0,1\}^d}$: that is,

$$d_{\text{DCS}}((x_\omega)_\omega, (y_\omega)_\omega) = \max_{\omega \in \{0,1\}^d} |x_\omega - y_\omega|.$$

Lemma 8.13 (Bounding the number of discrete Cantor sets). *Let $K \in \text{Int}(\mathbb{R})$ with $L := \mathcal{L}^1(K)$, and fix gap upper bounds $D = (D_1 \geq D_2 \geq \dots \geq D_d)$. Suppose $\delta \leq D_d/10, L/10$. Then*

$$\text{cov}((\text{DCS}_{d,D}(K), d_{\text{DCS}}), \delta) \leq \frac{2L}{\delta} \left(\frac{2D_1}{\delta}\right) \left(\frac{2D_2}{\delta}\right)^2 \dots \left(\frac{2D_d}{\delta}\right)^{2^{d-1}}.$$

Proof. The desired inequality is invariant under re-scaling \mathbb{R} , so we may simply assume that $\delta = 2$ and that $D_d, L \geq 20$.

Let $\Phi : \mathbb{R} \rightarrow \mathbb{Z}$ be the discretization map

$$\Phi(x) := \lfloor x \rfloor.$$

Clearly if $(x_\omega)_\omega, (y_\omega)_\omega \in \text{DCS}_{d,D}(K)$ and $(\Phi(x_\omega))_\omega = (\Phi(y_\omega))_\omega$ then

$$d_{\text{DCS}}((x_\omega)_\omega, (y_\omega)_\omega) < 2.$$

It therefore suffices bound the cardinality of the set $\Phi^{\times \{0,1\}^d}(\text{DCS}_{d,D}(K))$.

Let \overline{K} be a closed interval with end-points in \mathbb{Z} that contains K and has length at most $4L/3$. Using that $D_d, L \geq 20$, one sees that

$$\Phi^{\times \{0,1\}^d}(\text{DCS}_{d,D}(K)) \subseteq \text{DCS}_{d,4D/3}(\overline{K}) \cap \mathbb{Z}^{\{0,1\}^d},$$

so it suffices to bound the cardinality of this right-hand set by the desired product.

In the base case, $d = 1$, the set $\text{DCS}_{d,4D/3}(\overline{K}) \cap \mathbb{Z}^{\{0,1\}^d}$ just consists of pairs (x_0, x_1) in $\overline{K} \cap \mathbb{Z}$ separated by distance at most $4D_1/3$, and there are at most $(2L)(2D_1)$ of these.

Now, for the recursion clause, suppose the result is known for all depths less than some $d \geq 2$, and consider a discrete Cantor set $(x_\omega)_\omega \in \text{DCS}_{d,4D/3}(\overline{K}) \cap \mathbb{Z}^{\{0,1\}^d}$. It may be identified with the pair of depth- $(d-1)$ discrete Cantor sets

$$(x_{0\omega})_{\omega \in \{0,1\}^{d-1}} \quad \text{and} \quad (x_{1\omega})_{\omega \in \{0,1\}^{d-1}}.$$

Let

$$K_i := \min_{\omega \in \{0,1\}^{d-1}} x_{i\omega} + [0, 4D_2/3] \quad \text{for } i = 0, 1,$$

and let $D' := (D_2, \dots, D_d)$. In view of the gap upper bounds, the above two depth- $(d-1)$ discrete Cantor sets are members of

$$\text{DCS}_{d-1,4D'/3}(K_0) \cap \mathbb{Z} \quad \text{and} \quad \text{DCS}_{d-1,4D'/3}(K_1) \cap \mathbb{Z},$$

respectively.

Therefore the cardinality of $\text{DCS}_{d,4D/3}(\overline{K}) \cap \mathbb{Z}$ is bounded by the number of possible choices of K_1 and K_2 , multiplied by the square of $|\text{DCS}_{d-1,4D'/3}([0, 4D_2/3])|$. By the base-case argument and the inductive hypothesis, this is bounded by

$$\begin{aligned} (2L)(2D_1)|\text{DCS}_{d-1,2D'}([0, 4D_2/3])|^2 &\leq (2L)(2D_1)((2D_2)(2D_3)^2 \cdots (2D_d)^{2^{d-2}})^2 \\ &\leq (2L)(2D_1)(2D_2)^2(2D_3)^4 \cdots (2D_d)^{2^{d-1}}, \end{aligned}$$

as required. \square

Definition 8.14 (Discrete Cantor families). A **discrete Cantor family of depth d** is an indexed family $(K_\omega)_{\omega \in \{0,1\}^d}$ of pairwise-disjoint members of $\text{Int}(\mathbb{R})$. It has **gap upper bounds** $D_1 \geq \dots \geq D_d$ if

$$\text{diam}(K_\omega \cup K_{\omega'}) \leq D_{i+1} \quad \text{whenever } i < d \text{ and } \omega_j = \omega'_j \forall j \leq i.$$

If $\mathfrak{K} = (K_\omega)_{\omega \in \{0,1\}^d}$ is such a discrete Cantor family, then its **domain** is

$$\text{dom}(\mathfrak{K}) := \bigcup_{\omega \in \{0,1\}^d} K_\omega$$

(the reason for this terminology will become clear later).

We will let $\text{DCF}_{d,D}(K) \subseteq \text{Int}(K)^{\{0,1\}^d}$ denote the collection of all discrete Cantor families of depth d , contained in K , and having gap upper bounds given by D .

It is clear that if $(K_\omega)_\omega \in \text{DCF}_{d,D}(K)$, then

$$(\min K_\omega)_\omega \quad \text{and} \quad (\max K_\omega)_\omega \in \text{DCS}_{d,D}(K).$$

Similarly to d_{DCS} , we will endow $\text{DCF}_{d,D}(K)$ with the metric

$$\begin{aligned} d_{\text{DCF}}((K_\omega)_\omega, (K'_\omega)_\omega) &:= \max_{\omega \in \{0,1\}^d} d_{\text{Hdf}}(K_\omega, K'_\omega) \\ &= \max \{d_{\text{DCS}}((\min K_\omega)_\omega, (\min K'_\omega)_\omega), d_{\text{DCS}}((\max K_\omega)_\omega, (\max K'_\omega)_\omega)\}, \end{aligned}$$

where d_{Hdf} is the classical Hausdorff metric on the space of nonempty compact subsets. Lemma 8.13 immediately gives the following.

Corollary 8.15 (Bounding the number of discrete Cantor families). *Let $K \in \text{Int}(\mathbb{R})$ with $L := \mathcal{L}^1(K)$, and fix gap upper bounds $D = (D_1 \geq D_2 \geq \dots \geq D_d)$. Suppose $\delta \leq D_d/10, L/10$. Then*

$$\text{cov}((\text{DCF}_{d,D}(K), d_{\text{DCF}}), \delta) \leq \left(\frac{2L}{\delta} \left(\frac{2D_1}{\delta}\right) \left(\frac{2D_2}{\delta}\right)^2 \dots \left(\frac{2D_d}{\delta}\right)^{2^{d-1}}\right)^2.$$

□

We will also need to know how certain discrete Cantor families relate to the filtration \mathcal{D}_\bullet introduced previously. If $1 \leq r \leq d$ and $(K_\omega)_{\omega \in \{0,1\}^{d-r}}$ is a discrete Cantor family, then it is **adapted** to $(\mathcal{D}_d, \mathcal{D}_{d-1}, \dots, \mathcal{D}_r)$ if

- each $K_\omega \in \mathcal{D}_r$,
- if $s \in [0; d-r]$ and $\omega, \omega' \in \{0,1\}^{d-r}$ satisfy $\omega_i = \omega'_i$ for all $i \in [1; s]$, then

$$\mathcal{D}_{d-s}(K_\omega) = \mathcal{D}_{d-s}(K_{\omega'})$$

(including the case $s = 0$, when the assumption is vacuous),

- but if $\omega_s \neq \omega'_{s'}$ for some $s \in [1; d-r]$, then

$$\mathcal{D}_{d-s}(K_\omega) \neq \mathcal{D}_{d-s}(K_{\omega'}).$$

We now turn to the main result of this subsection, which provides discrete Cantor families on which a given cocycle-trajectory is (approximately) injective.

Proposition 8.16 (Finding a good discrete Cantor set for a good trajectory). *Let $1 \leq r < d$, let $Q \in \mathcal{D}_d$, let $y \in Y$, and suppose that $\mathcal{J} \subseteq \mathcal{D}_r \cap Q$ is a family of intervals such that*

$$|\mathcal{J}| \geq (1 - \kappa_{r,d})N_d/N_r$$

and

$$\bigcup \mathcal{J} \subseteq \bigcap_{s=r+1}^d H_{s,Q}^{\text{mndr}}(y).$$

Then there is a discrete Cantor family $(Q_\omega)_{\omega \in \{0,1\}^{d-r}}$ contained in \mathcal{J} and adapted to $(\mathcal{D}_d, \dots, \mathcal{D}_r)$ such that the images $B_\ell(\sigma_{Q_\omega}^y)$, $\omega \in \{0,1\}^{d-r}$, are also pairwise-disjoint.

Proof. This is proved by induction on d .

Base clause. When $d = r + 1$, our assumptions are

$$|\mathcal{J}| \geq (1 - \alpha_{r+1})L_{r+1}$$

and

$$\bigcup \mathcal{J} \subseteq H_{r+1,Q}^{\text{mndr}}(y).$$

The first of these implies that $\mathcal{J} \neq \emptyset$, and hence the second implies that also $H_{r+1,Q}^{\text{mndr}}(y) \neq \emptyset$. However, this is possible only if $H_{r+1,Q}^{\text{mndr}}(y) = Q$, and hence $S^{iN_{r+1}}y \in Y_{r+1,0}^{\text{mndr}}$ where $Q = [iN_{r+1}; (i+1)N_{r+1})$. Since $1 - \alpha_{r+1} \geq \alpha_{r+1}$, the definition of $Y_{r+1,0}^{\text{mndr}}$ gives two (necessarily disjoint) intervals $Q_0, Q_1 \in \mathcal{J}$ for which $B_\ell(\sigma_{Q_0}^y)$ and $B_\ell(\sigma_{Q_1}^y)$ are also disjoint.

Recursion clause. Now suppose $d > r + 1$, and that the result is already known at scales up to $d - 1$. Let \mathcal{J} satisfy the two assumptions, and let

$$\mathcal{J}_R := \{K \in \mathcal{J} \mid K \subseteq R\}$$

for each $R \in \mathcal{D}_{d-1} \cap Q$.

Our first assumption about \mathcal{J} gives

$$|(\mathcal{D}_r \cap Q) \setminus \mathcal{J}| = \sum_{R \in \mathcal{D}_{d-1} \cap Q} |(\mathcal{D}_r \cap R) \setminus \mathcal{J}_R| \leq \kappa_{r,d} N_d / N_r,$$

and so Markov's Inequality implies that the set

$$\mathcal{J}' := \{R \in \mathcal{D}_{d-1} \cap Q \mid |\mathcal{J}_R| \geq (1 - \kappa_{r,d-1})N_{d-1}/N_r\}$$

has cardinality at least $\alpha_d L_d$.

Our second assumption about \mathcal{J} requires that $H_{d,Q}^{\text{mndr}}(y) \neq \emptyset$, and hence instead $H_{d,Q}^{\text{mndr}}(y) = Q$. We may therefore apply the base-clause argument to the family $\mathcal{J}' \subseteq \mathcal{D}_{d-1} \cap Q$ to obtain a pair of intervals $Q_0, Q_1 \in \mathcal{J}'$ such that

$$B_\ell(\sigma_{Q_0}^y) \cap B_\ell(\sigma_{Q_1}^y) = \emptyset.$$

On the other hand, by the definition of \mathcal{J}' , for each $i \in \{1, 2\}$ we have

$$|\mathcal{J}_{Q_i}| \geq (1 - \kappa_{r,d-1})N_{d-1}/N_r,$$

and also

$$\begin{aligned} \bigcup \mathcal{J}_{Q_i} &= Q_i \cap \left(\bigcup \mathcal{J} \right) \subseteq Q_i \cap \left(\bigcap_{s=r+1}^d H_{s,Q}^{\text{mndr}}(y) \right) \\ &\subseteq \bigcap_{s=r+1}^{d-1} (Q_i \cap H_{s,Q}^{\text{mndr}}(y)) = \bigcap_{s=r+1}^{d-1} H_{s,Q_i}^{\text{mndr}}(y). \end{aligned}$$

We may therefore apply the inductive hypothesis to each of \mathcal{I}_{Q_0} and \mathcal{I}_{Q_1} to obtain Cantor families $(Q_{0\omega})_{\omega \in \{0,1\}^{d-r-1}}$ and $(Q_{1\omega})_{\omega \in \{0,1\}^{d-r-1}}$ inside them with the asserted properties. Assembling these into a single family shows that the induction continues, and hence completes the proof. \square

Remark 8.17. *Proposition 8.16 is a finitary cousin of the classical result that if $C \subseteq [0, 1]$ has Hausdorff dimension at most $\frac{1}{4}$, then a.e. Brownian sample path is injective on C : see Section 16.6 in Kahane [Kah85], up to Theorem 6 of that section. An unusual feature of the present setting is that our partitions of \mathbb{Z} are increasingly coarse, and each cell of \mathcal{D}_d contains roughly d^2 cells of \mathcal{D}_{d-1} , so the ‘index’ of \mathcal{D}_{d-1} in \mathcal{D}_d tends to ∞ with d . This is different from the classical analysis of self-crossing for Brownian motion on $[0, 1]$, which is easiest using simply dyadic partitions.* \triangleleft

8.6 Approximate covering with discrete Cantor families

Let (Y, σ) , φ and ℓ be as before.

Proposition 8.18. *For every $\beta, \eta > 0$ there are $M < \infty$, $r_0 \in \mathbb{N}$ and a family of subsets $Y_{r,d}^{\text{good}} \subseteq Y$, $d > r \geq r_0$, satisfying $\nu(Y_{r,d}^{\text{good}}) > 1 - \beta$ and such that the following holds. For every $r \geq r_0$ there is a $\delta > 0$ such that if $d > r$, $y \in Y_{r,d}^{\text{good}}$ and $P \subseteq [0; N_d]$ with $|P| > (1 - \delta)N_d$ then there is a collection \mathcal{G} of discrete Cantor families of depth $d - r$, all contained in P and subordinate to $(\mathcal{D}_d, \mathcal{D}_{d-1}, \dots, \mathcal{D}_{r+1})$, for which the following hold:*

- 1) (images are not too short) for every $(Q_\omega)_\omega \in \mathcal{G}$ and all $\omega \in \{0, 1\}^{d-r}$, one has

$$[\sigma_{\min Q_\omega}^y - N_r^{1/3}, \sigma_{\min Q_\omega}^y + N_r^{1/3}] \subseteq B_\ell(\sigma_{Q_\omega}^y);$$

- 2) (well-separated images) for every $(Q_\omega)_\omega \in \mathcal{G}$, the images $B_\ell(\sigma_{Q_\omega}^y)$ for $\omega \in \{0, 1\}^{d-r}$ are pairwise disjoint (beware that this is not asserting any disjointness among images from distinct members of \mathcal{G});

- 3) (cocycle-range is mostly covered) one has

$$\mathcal{L}^1\left(B_\ell(\sigma_{[0; N_d]}^y) \setminus \bigcup_{(Q_\omega)_\omega \in \mathcal{G}} \bigcup_{\omega \in \{0, 1\}^{d-r}} B_\ell(\sigma_{Q_\omega}^y)\right) \leq \eta \mathcal{L}^1(B_\ell(\sigma_{[0; N_d]}^y));$$

- 4) (cocycle-range is covered fairly efficiently) one has

$$\sum_{(Q_\omega)_\omega \in \mathcal{G}} \sum_{\omega \in \{0, 1\}^{d-r}} \mathcal{L}^1(B_\ell(\sigma_{Q_\omega}^y)) \leq M \mathcal{L}^1(B_\ell(\sigma_{[0; N_d]}^y)).$$

It seems worth emphasizing here that the final choice of δ must be allowed to depend on $r \geq r_0$.

Proof. Step 1: Choice of parameters and sets. We start with the selection of M and r_0 , along with some auxiliary parameters that will be used during the proof. We will then construct the sets $Y_{r,d}^{\text{good}}$.

The parameters are given by the following choices (which will be motivated during the course of the proof):

C1) Let $\varepsilon_1 := \eta/2$, and choose $r_{0,1} \in \mathbb{N}$ and $M_1 < \infty$ according to the convergence (28) so that $\nu(Y_{[0;N_d),M_1,\varepsilon_1}^{\text{smooth}}) > 1 - \beta/2$ for all $d > r_{0,1}$.

C2) Let $\delta_1 := \eta/4M_1$.

C3) Choose $r_{0,2}$ so large that

$$\kappa_{r_{0,2},\infty} > \max \left\{ 1 - \frac{\eta(1 - \delta_1)}{8M_1}, \delta_1 \right\}.$$

This implies that also

$$1 - \kappa_{r,d} < \frac{\eta(1 - \delta_1)}{8M_1} \quad \text{and} \quad \delta_1 < \kappa_{r,d} \quad \text{whenever } d > r \geq r_{0,2}.$$

C4) Let $\varepsilon_2 := \eta/16M_1$, and let $r_{0,3} \in \mathbb{N}$ and $M_2 < \infty$ be given by Corollary 8.11 so that for every $d > r \geq r_0$ the set

$$Y_{r,d}^{\text{good},1} := \left\{ y \mid \left| H_{r,[0;N_d)}^{\text{spread}}(y) \cap H_{r,[0;N_d),M_2,\varepsilon_2}^{\text{smooth}}(y) \cap \bigcap_{s=r+1}^d H_{s,[0;N_d)}^{\text{mndr}}(y) \right| \geq (1 - \delta_1/2)N_d \right\}$$

has

$$\nu(Y_{r,d}^{\text{good},1}) > 1 - \beta/2.$$

C5) Finally, let $M := 16M_1M_2/\eta$ and $r_0 := \max\{r_{0,1}, r_{0,2}, r_{0,3}\}$.

Now, for each $d > r$, let

$$Y_{r,d}^{\text{good}} := Y_{[0;N_d),M_1,\varepsilon_1}^{\text{smooth}} \cap Y_{r,d}^{\text{good},1},$$

where the second of these right-hand sets was introduced in choice (C4) above.

Choices (C1) and (C4) together imply that

$$\nu(Y_{r,d}^{\text{good}}) > 1 - \beta.$$

It remains to prove that these choices give the desired consequences.

Step 2: Finding a large family of good intervals. Now fix $d > r \geq r_0$, and let

$$\delta := \frac{\delta_1}{2N_r}.$$

We will complete the proof with this choice of δ . Thus, suppose that $y \in Y_{r,d}^{\text{good}}$ and that $P \subseteq [0; N_d)$ with $|P| > (1 - \delta)N_d$.

Consider the family of intervals

$$\mathcal{J}_0 := \{Q \in \mathcal{D}_r \mid Q \subseteq P\}.$$

For this family, one has

$$\begin{aligned} |(\mathcal{D}_r \cap [0; N_d)) \setminus \mathcal{J}_0| &\leq \sum_{Q \in (\mathcal{D}_r \cap [0; N_d)) \setminus \mathcal{J}_0} |Q \setminus P| \\ &\leq \sum_{Q \in \mathcal{D}_r \cap [0; N_d)} |Q \setminus P| = |[0; N_d) \setminus P| < \delta N_d, \end{aligned}$$

and hence

$$|\mathcal{J}_0| > (1 - \delta N_r)N_d/N_r = (1 - \delta_1/2)N_d/N_r.$$

Combining this bound with the fact that $y \in Y_{r,d}^{\text{good}} \subseteq Y_{r,d}^{\text{good},1}$, it follows that the family

$$\mathcal{J} := \left\{ Q \in \mathcal{D}_r \mid Q \subseteq P \cap H_{r,[0;N_d)}^{\text{spread}}(y) \cap H_{r,[0;N_d),M_2,\varepsilon_2}^{\text{smooth}}(y) \cap \bigcap_{s=r+1}^d H_{s,[0;N_d)}^{\text{mndr}}(y) \right\}$$

has

$$|\mathcal{J}| \geq (1 - \delta_1)N_d/N_r,$$

and this is greater than $(1 - \kappa_{r,d})N_d/N_r$ by choice (C3).

Step 3: Constructing the Cantor families Having introduced \mathcal{J} , Proposition 8.16 gives the ability to find a discrete Cantor family inside \mathcal{J} on which the cocycle-trajectories σ^y is approximately injective. It only remains to show how a careful repeated appeal to Proposition 8.16 can produce a whole collection of discrete Cantor families with the desired properties. This is achieved by the following recursion.

Base step. First, with \mathcal{J} as above, and in view of choice (C3), Proposition 8.16 gives a discrete Cantor family $\mathfrak{Q}_1 := (Q_\omega)_{\omega \in \{0,1\}^{d-r}}$ contained in \mathcal{J} such that

i) the intervals $B_\ell(\sigma_{Q_\omega}^y)$ for $\omega \in \{0, 1\}^{d-r}$ are pairwise disjoint.

Moreover, since $\bigcup \mathcal{J} \subseteq H_{r, [0; N_d]}^{\text{spread}}(y)$, we also have

ii) $[\sigma_{\min Q_\omega}^y - N_r^{1/3}, \sigma_{\min Q_\omega}^y + N_r^{1/3}] \subseteq B_\ell(\sigma_{Q_\omega}^y)$ for every ω .

Recursion step. Now suppose that discrete Cantor families $\Omega_1, \dots, \Omega_m$ in \mathcal{J} have already been constructed for some $m \geq 1$. If

$$\mathcal{L}^1\left(B_\ell(\sigma_{[0; N_d]}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y)\right) \leq \eta \mathcal{L}^1(B_\ell(\sigma_{[0; N_d]}^y)), \quad (31)$$

then stop the recursion, and let $\mathcal{G} := \{\Omega_1, \dots, \Omega_m\}$. In that case the construction is finished. Otherwise, suppose the opposite inequality of (31). Since $y \in Y_{r,d}^{\text{good}} \subseteq Y_{[0; N_d], M_1, \varepsilon_1}^{\text{smooth}}$, the opposite of (31) implies that

$$(\varphi \star \gamma_{[0; N_d]}^y) \left(B_\ell(\sigma_{[0; N_d]}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) > \frac{\eta - \varepsilon_1}{M_1} \geq \frac{\eta}{2M_1}. \quad (32)$$

Since

$$\varphi \star \gamma_{[0; N_d]}^y = \frac{N_r}{N_d} \sum_{Q \in \mathcal{D}_r \cap [0; N_d]} \varphi \star \gamma_Q^y \ll_{1, \delta_1} \frac{1}{|\mathcal{J}|} \sum_{Q \in \mathcal{J}} \varphi \star \gamma_Q^y,$$

inequality (32) gives that

$$\frac{1}{|\mathcal{J}|} \sum_{Q \in \mathcal{J}} (\varphi \star \gamma_Q^y) \left(B_\ell(\sigma_{[0; N_d]}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) > \frac{\eta}{2M_1} - \delta_1 > \frac{\eta}{4M_1},$$

by the choice of δ_1 in (C2). Therefore, letting

$$\mathcal{J}' := \left\{ Q \in \mathcal{J} \mid (\varphi \star \gamma_Q^y) \left(B_\ell(\sigma_{[0; N_d]}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) > \frac{\eta}{8M_1} \right\},$$

another appeal to Markov's Inequality implies that

$$|\mathcal{J}'| \geq \frac{\eta}{8M_1} |\mathcal{J}| \geq \frac{\eta(1 - \delta_1)}{8M_1} N_d / N_r,$$

and this is greater than $(1 - \kappa_{r,d})N_d/N_r$ by our choice in (C3), because $r \geq r_{0,2}$. We may therefore apply Proposition 8.16 again to obtain a discrete Cantor family $\Omega_{m+1} := (Q'_\omega)_{\omega \in \{0,1\}^{d-r}}$ in \mathcal{J}' satisfying the same properties (i) and (ii) as in the base step.

Observe that

$$(\varphi \star \gamma_{Q'_\omega}^y)(B_\ell(\sigma_{Q'_\omega}^y)) = 1 \quad \forall \omega \in \{0, 1\}^{d-r}.$$

Combined with the defining property of \mathcal{J}' , this gives

$$(\varphi \star \gamma_{Q'_\omega}^y) \left(B_\ell(\sigma_{Q'_\omega}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) > \frac{\eta}{8M_1}.$$

Since

$$\bigcup \mathcal{J}' \subseteq \bigcup \mathcal{J} \subseteq H_{r, [0; N_d], M_2, \varepsilon_2}^{\text{smooth}}(y),$$

this now implies that

$$\begin{aligned} \mathbb{U}_{B_\ell(\sigma_{Q'_\omega}^y)} \left(B_\ell(\sigma_{Q'_\omega}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) &> \frac{\frac{\eta}{8M_1} - \varepsilon_2}{M_2} > \frac{\eta}{16M_1M_2} = \frac{1}{M}. \\ \implies \mathcal{L}^1 \left(B_\ell(\sigma_{Q'_\omega}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) &\geq \frac{1}{M} \mathcal{L}^1(B_\ell(\sigma_{Q'_\omega}^y)). \end{aligned}$$

Therefore

$$\begin{aligned} &\mathcal{L}^1 \left(\bigcup_{s \leq m+1} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) \\ &= \sum_{\omega \in \{0, 1\}^{d-r}} \mathcal{L}^1 \left(B_\ell(\sigma_{Q'_\omega}^y) \setminus \bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \right) \\ &\geq \frac{1}{M} \sum_{\omega \in \{0, 1\}^{d-r}} \mathcal{L}^1(B_\ell(\sigma_{Q'_\omega}^y)) = \frac{1}{M} \mathcal{L}^1(B_\ell(\sigma_{\text{dom}(\Omega_{m+1})}^y)). \quad (33) \end{aligned}$$

Now continue the recursion.

Step 4: Completion of the proof. Since all quantities are finite, the algorithm described above must end after finitely many steps in some choice of $\mathcal{G} = \{\Omega_1, \dots, \Omega_m\}$. When it does so:

- conclusion (1) follows because all the chosen intervals were contained in $H_{r, [0; N_d]}^{\text{spread}}(y)$;
- conclusion (2) follows from the use of Proposition 8.16 at each step of the construction;
- conclusion (3) holds because it was the condition for the algorithm to terminate;

- conclusion (4) holds by an iterated appeal to inequality (33):

$$\begin{aligned}
& \sum_{s=1}^m \mathcal{L}^1(B_\ell(\sigma_{\text{dom}(\Omega_s)}^y)) \\
& \leq M \sum_{s=1}^m \mathcal{L}^1\left(B_\ell(\sigma_{\text{dom}(\Omega_s)}^y) \setminus \bigcup_{s'=1}^{s-1} B_\ell(\sigma_{\text{dom}(\Omega_{s'})}^y)\right) \\
& = M \mathcal{L}^1\left(\bigcup_{s \leq m} B_\ell(\sigma_{\text{dom}(\Omega_s)}^y)\right) \leq M \mathcal{L}^1(B_\ell(\sigma_{[0;N_d]}^y)).
\end{aligned}$$

□

9 The lower bound, completion of the proof

This section proves the lower bound in Theorem 5.23, and hence completes the proof of that theorem. It rests on a result showing that if y and y' satisfy some constraints in terms of $\sigma_{[0;N]}^y$ and $\sigma_{[0;N]}^{y'}$, and if $(y, x) \approx (y', x')$ according to $d_{[0;N]}^{\mathbf{Y} \times \sigma \mathbf{X}}$, then x and x' must themselves be ‘similar’, in a sense involving discrete Cantor sets.

This notion of similarity is quite cumbersome. I suspect this is a defect of the proof method, in that a much tighter (and more easily-stated) notion of similarity for sceneries also holds in our setting, as discussed in Subsection 4.3, but I have not been able to prove this.

It should be stressed that the structural relation between sceneries that we deduce is already implicitly at the heart of Kalikow’s argument in [Kal82], as well as its various sequels [Rud88, dHS97]. However, our formulation is superficially quite different from Kalikow’s. He introduces certain hierarchically-defined events in the space of walk-scenery pairs, involving longer and longer time-scales, and then uses a recursion to show that the probabilities of all these events remain close to 1. We re-interpret each of these as an event involving a ‘tree-like’ embedding of a discrete Cantor set into the domain of the scenery: this use of discrete Cantor sets converts Kalikow’s hierarchy of properties into a single geometric structure. In addition to making the relevant properties easier to visualize, this affords new ways of using Kalikow’s estimates: we will ultimately need to work with whole families of these discrete Cantor sets, whereas Kalikow needs to apply his hierarchical probability-estimate only once.

In the following, \mathbf{Y} , \mathbf{X} , σ , φ and ℓ will continue to be as in the previous section. We also continue to assume that $\text{diam}(X, d^X) \leq 1$.

9.1 Reduction to estimates for single bi-neighbourhoods

The key to the lower bound in Theorem 5.23 will be that, once d is large enough and for a suitable choice of the subset $U \subseteq Y \times X$ appearing in Definition 5.11, all bi-neighbourhoods in $(U, d_{[-N_d, 0]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}, d_{[0; N_d]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}, \nu \otimes \mu)$ are ‘small’ in a certain sense. This will then force one to use many of them to cover a positive proportion of U . This ‘smallness’ of the bi-neighbourhoods is in terms of the metric on the sceneries alone. Let $\pi_X : Y \times X \rightarrow X$ denote the coordinate projection, and let $\pi_X^* d^X$ denote the pullback of d^X to a pseudometric on $Y \times X$, and similarly for other metrics on X .

The following piece of notation will now be quite useful: given $(B, B') \in C_0(0, 1] \times C_0(0, 1]$, let

$$\text{aspect}(B, B') := \min \left\{ \frac{\mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]})}{\mathcal{L}^1(B_{[0,1]})}, \frac{\mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]})}{\mathcal{L}^1(B'_{[0,1]})} \right\},$$

interpreting this as 0 if either B or B' is constant. For $(B, B') \sim \mathbf{W}_{[0,1]}^{\otimes 2}$, the random variable $\text{aspect}(B, B')$ is a.s. positive with a continuous distribution on $(0, 1]$.

Proposition 9.1 (Bounding the covering number of a bi-neighbourhood). *Let \mathbf{X} and \mathbf{Y} be as before. For any $\gamma, \psi, \varepsilon, \beta, \delta' > 0$ there exist $\eta, \delta > 0$ and a sequence of subsets $Y_d \subseteq Y$, $d \geq 1$, such that $\nu(Y_d) > 1 - \beta$ for all sufficiently large d , and such that the following holds. If*

$$B, B' \in C_0(0, 1] \text{ with } \text{aspect}(B, B') > \gamma \text{ and } \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) \geq \psi/2,$$

and if we set $I := \sqrt{N_d} B_{[0,1]}$, $J := \sqrt{N_d} B'_{[0,1]}$, and

$$U := \{y \in Y_d \mid \|\text{traj}_{-N_d}(\sigma^y) - B\|_{\infty} < \eta \text{ and } \|\text{traj}_{N_d}(\sigma^y) - B'\|_{\infty} < \eta\} \times X,$$

then

$$\text{cov} \left(\left(U \cap B_{\delta N_d}^{d_{[0; N_d]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}} (U \cap B_{\delta N_d}^{d_{[-N_d; 0]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}} (y, x)), \pi_X^* d_{I \cap J}^{\mathbf{X}, \infty} \right), \delta' \right) \leq \exp(\varepsilon \mathcal{L}^1(I \cap J))$$

for any $(y, x) \in U$.

Heuristically, this proposition asserts that, provided one looks only within the set U defined by suitable approximate trajectories, both $d_{[-N_d; 0]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}$ and $d_{[0; N_d]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}$ remember something about the scenery x according to the metric $d_{I \cap J}^{\mathbf{X}, \infty}$. On the

other hand, we will be able to arrange that U has measure bounded below by some $\kappa > 0$, and so it will follow that one needs roughly $\exp(h(\mathbf{X})|I \cap J \cap \mathbb{Z}|)$ of these bi-neighbourhoods to cover a sizable portion of U .

Proposition 9.1 could be formulated using the Hamming-like metric $d_{I \cap J}^{\mathbf{X}}$ in place of the Bowen-Dinaburg metric $d_{I \cap J}^{\mathbf{X}, \infty}$, but the latter choice turns out to give a slightly shorter proof.

Most of this section will be given to the proof of Proposition 9.1. However, let us first show how it implies Theorem 5.23. This will also use the following auxiliary lemma.

Lemma 9.2. *For every $\alpha \in (1, \infty)$ and $\varepsilon > 0$ there exists $\gamma > 0$ for which the following holds. For any $\eta > 0$ there is a $\kappa > 0$ such that, for all sufficiently large $N \in \mathbb{N}$,*

$$\forall \nu' \in \Pr Y \text{ such that } \|\mathrm{d}\nu'/\mathrm{d}\nu\|_\infty \leq \alpha, \exists B, B' \in C_0(0, 1] \text{ such that}$$

$$\mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) > \psi_{\mathrm{BM}}(\alpha) - \varepsilon,$$

$$\text{aspect}(B, B') > \gamma,$$

and

$$\nu' \{y \mid \|\text{traj}_{-N}(\sigma^y) - B\|_\infty < \eta \text{ and } \|\text{traj}_N(\sigma^y) - B'\|_\infty < \eta\} \geq \kappa.$$

Proof. Let $\psi := \psi_{\mathrm{BM}}(\alpha)$. Definition 5.21 gives

$$W_{[0,1]}^{\otimes 2} \{(B, B') \mid \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) \leq \psi_{\mathrm{BM}}(\alpha) - \varepsilon\} < 1/\alpha.$$

Since $\text{aspect}(B, B') > 0$ for $W_{[0,1]}^{\otimes 2}$ -a.e. (B, B') , we may now choose $\gamma > 0$ such that the closed set

$$K := \{(B, B') \mid \mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) \leq \psi_{\mathrm{BM}}(\alpha) - \varepsilon \text{ or } \text{aspect}(B, B') \leq \gamma\}$$

still has $W_{[0,1]}^{\otimes 2}(K) < 1/\alpha$.

Now suppose that $\eta > 0$. By the inner-regularity of $W_{[0,1]}^{\otimes 2}$ with respect to compact sets, there are finitely many open subsets $W_1, \dots, W_m \subseteq C_0(0, 1]^2 \setminus K$ such that

- $W_{[0,1]}^{\otimes 2}(W_1 \cup \dots \cup W_m) > 1 - 1/\alpha + \kappa_0$ for some $\kappa_0 > 0$, and
- each W_i has diameter less than η for the maximum of the metrics $\|\cdot\|_\infty$ on each coordinate in $C_0(0, 1]^2$.

Set $\kappa := \alpha\kappa_0/m$, and choose representative pairs $(B_i, B'_i) \in W_i$ for each $i \leq m$.

The Invariance Principle (Theorem 3.10) implies that

$$(\text{traj}_{-N}(\sigma^y), \text{traj}_N(\sigma^y)) \xrightarrow{\text{law}} (B, B') \sim W_{[0,1]}^{\otimes 2} \quad \text{as } N \longrightarrow \infty.$$

Therefore, since each W_i is open, the Portmanteau Theorem ([Kal02, Theorem 4.25]) gives

$$\nu\{(\text{traj}_{-N}(\sigma^y), \text{traj}_N(\sigma^y)) \in W_1 \cup \dots \cup W_m\} > 1 - 1/\alpha + \kappa_0$$

for all sufficiently large N .

Suppose N is large enough that this last inequality holds, and now consider some $\nu' \in \Pr Y$ with $\|\text{d}\nu'/\text{d}\nu\|_\infty \leq \alpha$. Then that last inequality gives

$$\begin{aligned} & \nu'\{(\text{traj}_{-N}(\sigma^y), \text{traj}_N(\sigma^y)) \notin W_1 \cup \dots \cup W_m\} \\ & \leq \alpha \nu\{(\text{traj}_{-N}(\sigma^y), \text{traj}_N(\sigma^y)) \notin W_1 \cup \dots \cup W_m\} < \alpha(1/\alpha - \kappa_0) = 1 - \alpha\kappa_0, \end{aligned}$$

so there is some $i \leq m$ for which

$$\nu'\{(\text{traj}_{-N}(\sigma^y), \text{traj}_N(\sigma^y)) \in W_i\} > \kappa.$$

Letting $(B, B') := (B_i, B'_i)$ for this choice of i completes the proof. \square

Proof of Theorem 5.23 from Proposition 9.1. The upper bound has already been obtained, so it remains to prove the lower bound. This is vacuous if $h(\mathbf{Y}) = 0$, so assume $0 < h(\mathbf{Y}) < \infty$. We now imagine playing as Max-er in the competition of Subsection 5.2.

Step 1: The choice of measure and subset. First, Min-er chooses some $\lambda' \in \Pr(Y \times X)$ such that $\|\text{d}\lambda'/\text{d}(\nu \otimes \mu)\|_\infty \leq \alpha$. Let $\nu' \in \Pr Y$ and $\mu' \in \Pr X$ be its marginals, so we know that also $\|\text{d}\nu'/\text{d}\nu\|_\infty \leq \alpha$.

Let $\psi := \psi_{\text{BM}}(\alpha)$. Fix $\varepsilon > 0$, and assume without loss of generality that $\varepsilon < \psi/4$. For this ε , we will show that there are choices of $\delta > 0$ and $\kappa > \kappa' > 0$, depending on ε but not on the particular measure λ' , such that for each sufficiently large d there is some $U \subseteq Y \times X$ with $\lambda'(U) \geq \kappa$ and

$$\text{bicov}_{\kappa'}((U, d_{[-N_d; 0]}^{\mathbf{Y} \times_\sigma \mathbf{X}}, d_{[0; N_d]}^{\mathbf{Y} \times_\sigma \mathbf{X}}, \lambda'), \delta N_d) \geq \exp((h(\mathbf{X}) - 3\varepsilon)(\psi - 2\varepsilon)\sqrt{N_d}).$$

Since ε is arbitrary this will complete the proof.

The parameters and subset are chosen in the following steps:

- Let $\delta' > 0$ be so small that $h(\mu, T, d^X, \delta') > h(\mathbf{X}) - \varepsilon$, as is possible by Proposition 3.5.

- By the continuity of ψ_{BM} (Lemma 5.22), choose $\alpha' > \alpha$ such that $\psi_{\text{BM}}(\alpha') > \psi_{\text{BM}}(\alpha) - \varepsilon$.
- Let $\gamma > 0$ be given by Lemma 9.2 for α' and ε .
- Choose β so small that $\frac{\alpha}{1-\beta\alpha} < \alpha'$. Given $\gamma, \psi, \varepsilon, \delta'$ and this β , now apply Proposition 9.1 to obtain $\eta, \delta > 0$ and the subsets $Y_d \subseteq Y$ having the properties listed there. Observe that

$$\nu'(Y \setminus Y_d) \leq \alpha \nu(Y \setminus Y_d) < \alpha\beta \implies \lambda'(Y_d \times X) = \nu'(Y_d) \geq 1 - \alpha\beta,$$

and hence

$$\left\| \frac{d\lambda'_{|Y_d \times X}}{d(\nu \otimes \mu)} \right\|_\infty \leq \frac{\alpha}{1 - \alpha\beta} < \alpha'$$

for all sufficiently large d .

- For our given ε and for the values of γ and η chosen above, and for d sufficiently large, now return to Lemma 9.2, applied to the measure $\nu'_{|Y_d}$, to obtain some $\kappa > 0$ such that, for all sufficiently large d , there are $B, B' \in C_0(0, 1]$ satisfying

$$\mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) > \psi_{\text{BM}}(\alpha') - \varepsilon > \psi - 2\varepsilon > \psi/2,$$

$$\text{aspect}(B, B') > \gamma,$$

and such that the set

$$U_1 := \{y \in Y_d \mid \|\text{traj}_{-N_d}(\sigma^y) - B\|_\infty < \eta \text{ and } \|\text{traj}_{N_d}(\sigma^y) - B'\|_\infty < \eta\}$$

has

$$\nu'(U_1) = \nu'(Y_d)\nu'_{|Y_d}(U_1) \geq (1 - \alpha\beta)\nu'_{|Y_d}(U_1) \geq \kappa.$$

Let $I := \sqrt{N_d}B_{[0,1]}$ and $J := \sqrt{N_d}B'_{[0,1]}$.

- Finally, let $\kappa' := \kappa/2$, and let $U := U_1 \times X$, so $\lambda'(U) = \nu'(U_1) \geq \kappa$ for all sufficiently large d .

Step 2: Bounding the bi-covering number. The conclusion of Proposition 9.1 now gives that

$$\text{cov}\left(\left(U \cap B_{\delta N_d}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}^{[0; N_d]}\right)(U \cap B_{\delta N_d}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}^{[-N_d; 0]}(y, x)), \pi_X^* d_{I \cap J}^{\mathbf{X}, \infty}\right), \delta'\right) \leq \exp(\varepsilon \mathcal{L}^1(I \cap J))$$

for all $(y, x) \in U$, for all sufficiently large d . On the other hand, if $V \subseteq U$ with $\lambda'(V) \geq \kappa'$, then also $(\nu \otimes \mu)(V) \geq \kappa'/\alpha$ and hence $\mu(\pi_X(V)) \geq \kappa'/\alpha$, a fixed positive constant. Therefore, provided d and hence $\mathcal{L}^1(I \cap J) \geq (\psi/2)\sqrt{N_d}$ are sufficiently large, Proposition 3.6 and the choice of δ' give

$$\begin{aligned} \text{cov}((V, \pi_X^* d_{I \cap J}^{\mathbf{X}, \infty}), \delta') &= \text{cov}((\pi_X(V), d_{I \cap J}^{\mathbf{X}, \infty}), \delta') \\ &\geq \text{cov}((\pi_X(V), d_{I \cap J}^{\mathbf{X}, \infty}), \delta' \mathcal{L}^1(I \cap J)) \\ &> \exp((h(\mathbf{X}) - 2\varepsilon) \mathcal{L}^1(I \cap J)). \end{aligned}$$

These two bounds together imply that, if d is sufficiently large, then the number of (δN_d) -bi-neighbourhoods needed to cover such a subset $V \subseteq U$ is at least

$$\frac{\exp((h(\mathbf{X}) - 2\varepsilon) \mathcal{L}^1(I \cap J))}{\exp(\varepsilon \mathcal{L}^1(I \cap J))} \geq \exp((h(\mathbf{X}) - 3\varepsilon)(\psi - 2\varepsilon)\sqrt{N_d}),$$

as required. \square

The rest of this section is occupied by the proof of Proposition 9.1.

9.2 Discrete Cantor matchings

First we need the following relative of Definition 8.12. Its importance will appear in the formulation of Proposition 9.4.

Definition 9.3 (Discrete Cantor matchings). *Let $d \in \mathbb{N}$. A **discrete Cantor matching of depth d** is a pair $(K_\omega, u_\omega)_{\omega \in \{0,1\}^d}$ in which $(K_\omega)_\omega$ is a discrete Cantor family and $(u_\omega)_\omega$ is a discrete Cantor set, both of depth d . It has **gap upper bounds** $D_1 \geq \dots \geq D_d$ if these are gap upper bounds for both this discrete Cantor family and this discrete Cantor set.*

*If $\mathfrak{M} = (K_\omega, u_\omega)_\omega$ is a discrete Cantor matching, then its **domain** is*

$$\text{dom}(\mathfrak{M}) := \bigcup_{\omega} K_\omega.$$

For fixed d , $D = (D_1 \geq \dots \geq D_d)$ and $J \in \text{Int}(\mathbb{R})$, the collection of depth- d discrete Cantor matchings contained in J and with gap upper bounds D will be denoted by $\text{DCM}_{d,D}(J)$.

Clearly $\text{DCM}_{d,D}(J)$ may be identified with $\text{DCF}_{d,D}(J) \times \text{DCS}_{d,D}(J)$. We endow $\text{DCM}_{d,D}(J)$ with the metric d_{DCM} given by the maximum of the metrics

d_{DCF} and d_{DCS} on these coordinate factors, and so Lemma 8.13 and Corollary 8.15 immediately imply

$$\begin{aligned} \text{cov}((\text{DCM}_{d,D}(J), d_{\text{DCM}}), \delta) \\ \leq \left(\frac{2\mathcal{L}^1(J)}{\delta} \left(\frac{2D_1}{\delta} \right) \left(\frac{2D_2}{\delta} \right)^2 \cdots \left(\frac{2D_d}{\delta} \right)^{2^{d-1}} \right)^3 \end{aligned} \quad (34)$$

provided $\delta < \mathcal{L}^1(J)/10, D_d/10$.

9.3 Similarity of sceneries from similarity of pairs

Let the parameters L_d, N_d, α_d and $\kappa_{r,d}$ be as introduced in Subsection 8.4.

Proposition 9.4. *Suppose that $\beta, \eta > 0$, and let $M < \infty, r_0 \in \mathbb{N}$, and the sets $Y_{r,d}^{\text{good}}$ for $d > r \geq r_0$ be as provided by Proposition 8.18 for this β and η .*

Then for any $r \geq r_0$ there is a $\delta > 0$ such that if

- $B \in C_0(0, 1]$ and $J := \sqrt{N_d}B_{[0,1]} + [-\eta\sqrt{N_d}, \eta\sqrt{N_d}]$,
- $(y, x), (y', x') \in Y_{r,d}^{\text{good}} \times X$ with

$$\|\text{traj}_{N_d}(\sigma^y) - B\|_\infty, \|\text{traj}_{N_d}(\sigma^{y'}) - B\|_\infty < \eta,$$

- and

$$d_{[0;N_d]}^{\mathbf{Y} \times_\sigma \mathbf{X}}((y, x), (y', x')) \leq \delta N_d, \quad (35)$$

then there is a tuple

$$\mathcal{F} \in \text{DCM}_{d-r, 4\ell(N_d \geq \dots \geq N_{r+1})}(J)^m \quad \text{for some } m \leq \frac{M\mathcal{L}^1(J)}{2^{d-r}N_r^{1/3}}$$

such that the following hold:

P1) (discrete Cantor sets are small) for every $(K_\omega, u_\omega)_\omega \in \mathcal{F}$ and every $\omega \in \{0, 1\}^{d-r}$, one has $|u_\omega| < 2\eta\sqrt{N_d}$;

P2) (range interval is mostly covered) one has

$$\mathcal{L}^1\left(J \setminus \bigcup_{\mathfrak{M} \in \mathcal{F}} \text{dom}(\mathfrak{M})\right) < \eta\mathcal{L}^1(J) + 4\eta\sqrt{N_d};$$

P3) (sceneries approximately agree across matchings) for every $(K_\omega, u_\omega)_\omega \in \mathcal{F}$ and $\omega \in \{0, 1\}^{d-r}$, one has

$$d^X(T^z x, T^{z+u_\omega} x') < \eta \quad \forall z \in K_\omega.$$

Remark 9.5. Beware that the assumptions of this proposition are symmetric between (y, x) and (y', x') , but the conclusions are not. \triangleleft

Proof. Step 1: Choice of parameters. Suppose that $r \geq r_0$, and let $\delta_1 > 0$ be the error tolerance given by Proposition 8.18 for this r .

Since the action $T : \mathbb{R} \curvearrowright X$ is jointly continuous, given our $\eta > 0$ and ℓ , there is some $\tilde{\eta} > 0$ such that for any $x, x' \in X$ one has

$$d^X(x, x') < 2\tilde{\eta} \implies d^X(T^z x, T^z x') < \eta \quad \forall z \in [-\ell, \ell].$$

Again by the continuity of this action, we may now choose $\varepsilon > 0$ such that

$$|z| \leq \varepsilon \implies d^X(x, T^z x) < \tilde{\eta} \quad \forall x \in X.$$

Next, given this ε , Corollary 7.2 gives some $\tilde{\delta} > 0$ such that for any discrete interval $[0; L) \subseteq \mathbb{Z}$ and any $x, x' \in Y$ one has

$$\max_{n \in [0; L)} d^Y(S^n y, S^n y') < \tilde{\delta} \implies \max_{n \in [0; L)} |\sigma_n^y - \sigma_n^{y'}| < \varepsilon.$$

Finally, let $\delta := \delta_1 \cdot \min\{\tilde{\delta}, \tilde{\eta}\}$, and assume (35) with this value of δ .

Step 2: Using Proposition 8.18. Consider the set

$$P := \{n \in [0; N_d) \mid d^Y(S^n y, S^n y') \leq \tilde{\delta} \text{ and } d^X(T^{\sigma_n^y} x, T^{\sigma_n^{y'}} x') \leq \tilde{\eta}\}.$$

By (35), our choice of δ and Markov's Inequality, one has

$$\delta N_d \geq |[0; N_d) \setminus P| \cdot \min\{\tilde{\delta}, \tilde{\eta}\} \implies |P| \geq (1 - \delta_1) N_d.$$

We may therefore subject P to an application of Proposition 8.18. Let \mathcal{G} be the collection of discrete Cantor families produced by that proposition, ordered so that $\mathcal{G} = ((Q_{s,\omega})_\omega)_{s=1}^m$ for some m , and let

$$\mathcal{F} := ((B_\ell(\sigma_{Q_{s,\omega}}^y), \sigma_{\min Q_{s,\omega}}^{y'} - \sigma_{\min Q_{s,\omega}}^y)_{\omega \in \{0,1\}^{d-r}})_{s=1}^m.$$

To see that each entry of \mathcal{F} is a member of $\text{DCM}_{d-r, 4\ell(N_d \geq \dots \geq N_{r+1})}(J)$, observe the following:

- property (2) of Proposition 8.18 gave that the image-neighbourhoods $B_\ell(\sigma_{Q_{s,\omega}}^y)$ are pairwise disjoint for distinct $\omega \in \{0, 1\}^{d-r}$, for each fixed s ;
- these neighbourhoods are all contained in $B_\ell(\sigma_{[0;N_d]}^y) \subseteq J$;
- and, since $\ell \geq \|\sigma\|_\infty$, for any bounded discrete interval $R \subseteq \mathbb{Z}$ and any $y \in Y$, the image σ_R^y must have diameter at most $\ell \mathcal{L}^1(R)$. It follows that $B_\ell(\sigma_R^y)$ is an interval of length at most $\ell \mathcal{L}^1(R) + 2\ell$. Therefore, for each $s \leq m$, the gap upper bounds $\ell(N_d + 2) \geq \dots \geq \ell(N_{r+1} + 2)$ hold for the discrete Cantor family $(B_\ell(\sigma_{Q_{s,\omega}}^y))_\omega$ because $(Q_\omega)_\omega$ was adapted to $(\mathcal{D}_d, \dots, \mathcal{D}_{r+1})$, and the gap upper bounds $2\ell(N_d + 2) \geq \dots \geq 2\ell(N_{r+1} + 2)$ hold for the discrete Cantor set $(\sigma'_{\min Q_{s,\omega}} - \sigma_{\min Q_{s,\omega}}^y)_\omega$ as it is a set of differences of elements of such discrete Cantor families. This conclusion is stronger than the gap upper bounds $4\ell N_d \geq \dots \geq 4\ell N_{r+1}$ of property (P3), because $N_{r+1} \geq N_{r_0+1} \geq 2$.

The desired upper bound on m holds because property (1) of Proposition 8.18 gives

$$\begin{aligned} \mathcal{L}^1(K_\omega) &\geq N_r^{1/3} \quad \forall (K_\omega, u_\omega)_\omega \in \mathcal{F} \text{ and } \omega \in \{0, 1\}^{d-r} \\ &\implies \mathcal{L}^1(\text{dom}(\mathfrak{M})) \geq 2^{d-r} N_r^{1/3} \quad \forall \mathfrak{M} \in \mathcal{F}, \end{aligned}$$

while property (4) of that proposition gives

$$\sum_{\text{dom}(\mathfrak{M}) \in \mathcal{F}} \mathcal{L}^1(\text{dom}(\mathfrak{M})) \leq M \mathcal{L}^1(B_\ell(\sigma_{[0;N_d]}^y)) \leq M \mathcal{L}^1(J).$$

Step 3: Verifying the remaining properties. It remains to prove (P1)–(P3).

Property (P1) holds because $\text{traj}_N(\sigma^y)$ and $\text{traj}_N(\sigma^{y'})$ are both close to B : for $t := \min Q_\omega / N_d$, those approximations give

$$|\sigma_{\min Q_\omega}^y - \sigma'_{\min Q_\omega}| = \sqrt{N_d} |\text{traj}_{N_d}(\sigma^y)(t) - \text{traj}_{N_d}(\sigma^{y'})(t)| < 2\eta \sqrt{N_d}.$$

Property (P2) results from property (3) of Proposition 8.18, combined with the facts that

$$B_\ell(\sigma_{[0;N_d]}^y) \subseteq J \quad \text{and} \quad \mathcal{L}^1(J \setminus B_\ell(\sigma_{[0;N_d]}^y)) < 4\eta \sqrt{N_d}.$$

Finally, property (P3) holds because for each $s \leq m$ and each $\omega \in \{0, 1\}^{d-r}$, we have that $Q_{s,\omega} \subseteq P$ by construction, and so the definition of P gives

$$d^Y(S^n y, S^n y') < \tilde{\delta} \quad \text{and} \quad d^X(T^{\sigma_n^y} x, T^{\sigma_n^{y'}} x') < \tilde{\eta} \quad \forall n \in Q_{s,\omega}. \quad (36)$$

For this s and each ω , now abbreviate $n_\omega := \min Q_{s,\omega}$ and $u_\omega := \sigma_{n_\omega}^{y'} - \sigma_{n_\omega}^y$. By the choice of $\tilde{\delta}$, the first inequality in (36) implies that

$$|\sigma_n^{y'} - \sigma_n^y - u_\omega| = |(\sigma_n^{y'} - \sigma_{n_\omega}^{y'}) - (\sigma_n^y - \sigma_{n_\omega}^y)| = |\sigma_{n-n_\omega}^{S^{n_\omega} y'} - \sigma_{n-n_\omega}^{S^{n_\omega} y}| < \varepsilon \quad \forall n \in Q_{s,\omega}.$$

Given this, and letting $y_1 := S^{n_\omega} y$, $y'_1 := S^{n_\omega} y'$, $x_1 := T^{\sigma_{n_\omega}^y} x$ and $x'_1 := T^{\sigma_{n_\omega}^{y'}} x'$, the second inequality in (36) may be re-written as

$$d^X(T^{\sigma_n^{y_1}} x_1, T^{\sigma_n^{y'_1}} x'_1) < \tilde{\eta} \quad \forall n \in [0; N_r).$$

Combining the above inequalities, and recalling the choice of ε , we now obtain

$$d^X(T^{\sigma_n^{y_1}} x_1, T^{\sigma_n^{y'_1}} x'_1) < 2\tilde{\eta} \quad \forall n \in [0; N_r),$$

and now by the choice of $\tilde{\eta}$ this implies that

$$d^X(T^z x, T^{z+u_\omega} x') < \eta \quad \forall z \in B_\ell(\sigma_{[0; N_r)}^{y_1}) = B_\ell(\sigma_{Q_{s,\omega}}^y).$$

□

We will retain the names (P1)–(P3) for the above properties throughout the rest of the paper. Note that, by duplicating some members of the resulting family \mathcal{F} , we may always assume that $m = \lfloor M \mathcal{L}^1(J) / 2^{d-r} N_r^{1/3} \rfloor$ without disrupting these other properties.

9.4 Bounding the covering number of discrete Cantor matchings

The next lemma is an elementary estimate which will lie at the heart of the competition between two different sources of entropy in the sequel.

Lemma 9.6. *There is some absolute constant $C_0 < \infty$ such that*

$$\sum_{s=0}^{d-r} 2^s (d-s) \log(d-s) \leq C_0 2^d \quad \text{whenever } d > r \geq 1.$$

Proof. Dividing the left-hand side by 2^d produces the sum

$$\sum_{s=0}^{d-r} 2^{-(d-s)} (d-s) \log(d-s) = \sum_{\ell=r}^d 2^{-\ell} \ell \log \ell \lesssim \sum_{\ell=r}^{\infty} 2^{-\ell} \cdot 2^{\ell/2} < \infty.$$

□

For any fixed d , $D = (D_1 \geq \dots \geq D_d)$ and $J \in \text{Int}(\mathbb{R})$, and for each $m \in \mathbb{N}$, let $d_{\text{DCM},m}$ be the metric on $(\text{DCM}_{d,D}(J))^m$ given as the maximum of the metric d_{DCM} on each of the m coordinates.

Lemma 9.7 (Bounding the number of discrete-Cantor-matching tuples). *For every $\delta, \varepsilon > 0$ there exists $r_1 \in \mathbb{N}$ such that if $d > r \geq r_1$, if $J \in \text{Int}(\mathbb{R})$ has length at most $4\ell N_d$, and if*

$$m := \left\lfloor \frac{M \mathcal{L}^1(J)}{2^{d-r} N_r^{1/3}} \right\rfloor,$$

then

$$\text{cov}\left(\left((\text{DCM}_{d-r, 4\ell(N_d \geq \dots \geq N_{r+1})}(J))^m, d_{\text{DCM},m}\right), \delta\right) \leq \exp(\varepsilon \mathcal{L}^1(J)) \quad \forall m \in \mathbb{N}.$$

Proof. By (34) and (29), there is an absolute constant $C < \infty$ such that this covering number is bounded by

$$\begin{aligned} & \left(\frac{8\ell N_d}{\delta} \left(\frac{8\ell N_d}{\delta} \right) \left(\frac{8\ell N_{d-1}}{\delta} \right)^2 \dots \left(\frac{8\ell N_{r+1}}{\delta} \right)^{2^{d-r-1}} \right)^{3m} \\ &= \left(\frac{8\ell}{\delta} \right)^{3m 2^{d-r}} (N_d \cdot N_d \cdot N_{d-1}^2 \cdots N_{r+1}^{2^{d-r-1}})^{3m} \\ &\leq \exp \left(Cm \left((d+1) \log(d+1) + 2d \log d + \dots + 2^{d-r-1} (r+2) \log(r+2) \right) \right. \\ &\quad \left. + 3m 2^{d-r} \log \frac{8\ell}{\delta} \right) \\ &= \exp \left(Cm \sum_{s=0}^{d-r-1} 2^s (d+1-s) \log(d+1-s) + 3m 2^{d-r} \log \frac{8\ell}{\delta} \right). \end{aligned}$$

Substituting for m , this is bounded by

$$\exp \left(\frac{C' \mathcal{L}^1(J)}{2^{d-r} N_r^{1/3}} \sum_{s=0}^{d-r-1} 2^s (d+1-s) \log(d+1-s) + C'' \frac{\mathcal{L}^1(J)}{N_r^{1/3}} \right),$$

where

$$C' := CM \quad \text{and} \quad C'' := 3M \log \frac{8\ell}{\delta},$$

neither of which depends on d or r . Letting C_0 be the constant from Lemma 9.6, the above expression is in turn bounded by

$$\exp \left(\frac{2C' C_0 \mathcal{L}^1(J)}{N_r^{1/3} 2^{-r}} + C'' \frac{\mathcal{L}^1(J)}{N_r^{1/3}} \right).$$

Another appeal to (29) implies that $N_r^{1/3} 2^{-r} \rightarrow \infty$ as $r \rightarrow \infty$, so the above is bounded by $\exp(\varepsilon \mathcal{L}^1(J))$ provided r was large enough, irrespective of the value of d . \square

9.5 Bounding covering numbers of bi-neighbourhoods

Lemma 9.8. *For any $\gamma', \zeta > 0$ there are $\eta, \delta''' \in (0, 1)$ for which the following holds. Suppose that $K \subseteq J \subseteq \mathbb{R}$ are compact intervals, both containing 0, such that $\mathcal{L}^1(K) \geq \gamma' \mathcal{L}^1(J)$. Suppose also that $d > r \geq 1$, that $x_1, x'_1, x_2, x'_2 \in X$, and that*

$$m \in \mathbb{N} \quad \text{and} \quad \mathcal{F}, \mathcal{G} \in \left(\text{DCM}_{d-r, 4\ell(N_d \geq \dots \geq N_{r+1})}(J) \right)^m$$

are such that

J, x_1, x'_1 and \mathcal{F} satisfy (P1)–(P3) for these values of r, d, η ,

J, x_2, x'_2 and \mathcal{G} satisfy (P1)–(P3) for these values of r, d, η ,

$$d_{\text{DCM}, m}(\mathcal{F}, \mathcal{G}) < \delta''', \tag{37}$$

and

$$d_K^{\mathbf{X}, \infty}(x'_1, x'_2) < \delta'''. \tag{38}$$

Then

$$d_K^{\mathbf{X}}(x_1, x_2) < \zeta(\mathcal{L}^1(K) + \sqrt{N_d}).$$

(The choice of the notation ‘ γ' ’ and ‘ δ''' ’ is for ease of reference later.)

Remarks 9.9. (1.) *It is very important here that we assume only proximity of x_1 and x_2 in $d_K^{\mathbf{X}, \infty}$, rather than $d_J^{\mathbf{X}}$, and then (of course) also conclude only that kind of proximity. On the other hand, it is also important that the input is an inequality for $d_K^{\mathbf{X}, \infty}$, whereas the output is only for $d_K^{\mathbf{X}}$; this difference will be taken into account later by an appeal to Lemma 3.7.*

(2.) *Note also that this lemma does not use the bound on the length m of the tuples \mathcal{F} and \mathcal{G} .* \triangleleft

Proof. First, using the joint continuity of T , choose $\delta_0 > 0$ so small that

$$|z| \leq 2\delta_0 \implies \max_{x \in X} d^X(x, T^z x) < \zeta/4.$$

Next, choose $\eta > 0$ and $\delta''' \in (0, \delta_0]$ both so small that

$$\eta/\gamma' + 2\eta + \delta''' < \zeta/2 \quad \text{and} \quad 8\eta < \zeta.$$

Now assume also that $d > r \geq 1$, and that $x_1, x'_1, x_2, x'_2, \mathcal{F}$ and \mathcal{G} satisfy the stated assumptions. For each $s \leq m$, let

$$\mathcal{F} = \left((K_{1,s,\omega}, u_{1,s,\omega})_{\omega \in \{0,1\}^{d-r}} \right)_{s=1}^m \quad \text{and} \quad \mathcal{G} = \left((K_{2,s,\omega}, u_{2,s,\omega})_{\omega \in \{0,1\}^{d-r}} \right)_{s=1}^m.$$

Let

$$D := \bigcup_{s=1}^m \bigcup_{\omega \in \{0,1\}^{d-r}} K_{1,s,\omega},$$

and let $K' \subseteq K$ be the closed subinterval with the same centre and length $\mathcal{L}^1(K) - 4\eta\sqrt{N_d}$ (understood as \emptyset if this value is negative). By property (P2),

$$\begin{aligned} \mathcal{L}^1(K \setminus (K' \cap D)) &\leq \mathcal{L}^1(K \setminus K') + \mathcal{L}^1(J \setminus D) \\ &\leq 4\eta\sqrt{N_d} + (\eta\mathcal{L}^1(J) + 4\eta\sqrt{N_d}) \leq 8\eta\sqrt{N_d} + (\eta/\gamma')\mathcal{L}^1(K). \end{aligned}$$

Now suppose that $t \in K' \cap D$. There are $s \leq m$ and $\omega \in \{0,1\}^{d-r}$ such that $t \in K_{1,s,\omega}$, and now the approximation (37) gives some $w \in (-\delta''', \delta''')$ such that $t + w \in K_{2,s,\omega}$. Let $u_i := u_{i,s,\omega}$ for $i = 1, 2$.

By the triangle inequality,

$$\begin{aligned} d^X(T^t x_1, T^t x_2) &\leq d^X(T^t x_1, T^{t+u_1} x'_1) + d^X(T^{t+u_1} x'_1, T^{t+u_1} x'_2) \\ &\quad + d^X(T^{t+u_1} x'_2, T^{t+w+u_2} x'_2) + d^X(T^{t+w+u_2} x'_2, T^{t+w} x_2) \\ &\quad + d^X(T^{t+w} x_2, T^t x_2). \end{aligned}$$

These five right-hand terms may now be bounded separately:

- property (P3) gives

$$d^X(T^t x_1, T^{t+u_1} x'_1) < \eta \quad \text{and} \quad d^X(T^{t+w+u_2} x'_2, T^{t+w} x_2) < \eta;$$

- since $t \in K''$ and property (P1) gives $|u_1| < 2\eta\sqrt{N_d}$, we still have $t + u_1 \in K'$, and so the approximation (38) gives

$$d^X(T^{t+u_1} x'_1, T^{t+u_1} x'_2) < \delta''';$$

- finally, our assumptions gave $|w| \leq \delta'''$ and the approximation (37) gives $|u_1 - u_2| < \delta$, so the choice of $\delta''' \leq \delta_0$ implies that

$$d^X(T^{t+u_1} x'_2, T^{t+w+u_2} x'_2) < \zeta/4 \quad \text{and} \quad d^X(T^{t+w} x_2, T^t x_2) < \zeta/4.$$

Putting these estimates together gives

$$d^X(T^t x_1, T^t x_2) < 2\eta + \delta + \zeta/2 \quad \forall t \in K' \cap D.$$

Integrating over $t \in K'$, this becomes

$$\begin{aligned} d_{K'}^{\mathbf{X}}(x_1, x_2) &\leq \int_{K \setminus (K' \cap D)} d^X(T^t x_1, T^t x_2) dt + \int_{K' \cap D} d^X(T^t x_1, T^t x_2) dt \\ &\leq \mathcal{L}^1(K \setminus (K' \cap D)) + (2\eta + \delta''' + \zeta/2) \mathcal{L}^1(K' \cap D) \\ &\leq (\eta/\gamma') \mathcal{L}^1(K) + 8\eta\sqrt{N_d} + (2\eta + \delta''' + \zeta/2) \mathcal{L}^1(K) \\ &= (\eta/\gamma' + 2\eta + \delta''' + \zeta/2) \mathcal{L}^1(K) + 8\eta\sqrt{N_d}, \end{aligned}$$

and this is less than $\zeta \mathcal{L}^1(K) + \zeta\sqrt{N_d}$ by the choice of η and δ''' . \square

Lemma 9.10. *For any $\gamma, \psi, \varepsilon, \beta, \delta' > 0$ there are $\eta, \delta > 0$ and a sequence of subsets $Y_d \subseteq Y$, $d \geq 1$, such that $\nu(Y_d) > 1 - \beta$ for all sufficiently large d , and such that the following holds. If*

$$B \in C_0(0, 1], J := \sqrt{N_d} B_{[0, 1]} \text{ and } K \in \text{Int}(J) \text{ satisfy both } \mathcal{L}^1(K) \geq (\psi/2)\sqrt{N_d} \text{ and } \mathcal{L}^1(K) \geq \gamma \mathcal{L}^1(J),$$

and if

$$U \subseteq \{y \in Y_d \mid \|\text{traj}_{N_d}(\sigma^y) - B\|_\infty < \eta\} \times X,$$

then

$$\text{cov}((U \cap B_{\delta N_d}^{d_{[0; N_d]}^{\mathbf{Y} \times_\sigma \mathbf{X}}}(C), \pi_X^* d_K^{\mathbf{X}, \infty}), \delta') \leq \exp(\varepsilon \mathcal{L}^1(K))$$

whenever $C \subseteq U$ has diameter at most δ according to the pseudometric $\pi_X^* d_K^{\mathbf{X}, \infty}$.

The analogous result holds when traj_{N_d} is replaced by traj_{-N_d} and $d_{[0; N_d]}^{\mathbf{Y} \times_\sigma \mathbf{X}}$ is replaced by $d_{[-N_d; 0]}^{\mathbf{Y} \times_\sigma \mathbf{X}}$.

Proof. It will be clear that the second assertion follows in the same way as the first, so we concentrate on that.

Step 1: Choosing the parameters.

- First recall from Lemma 3.7 that for our given $\delta' > 0$, there is some $\delta'' > 0$ such that for every $x \in X$ and $K \in \text{Int}(\mathbb{R})$ with $\mathcal{L}^1(K) \geq 1$ one has

$$\text{cov}((B_{\delta'' \mathcal{L}^1(K)}^{d_K^{\mathbf{X}}}(x), d_K^{\mathbf{X}, \infty}), \delta') < \exp(\varepsilon \mathcal{L}^1(K)/2).$$

- Next, choose $\gamma' := \gamma/(1+4/\psi)$, choose ζ so small that $\zeta(1+2/\psi) < \delta''$, and now implement Lemma 9.8 with this γ' and ζ to obtain some $\eta, \delta''' \in (0, 1)$ with the property described there.
- For the given value of β and for the η chosen above, now let $M < \infty$, $r_0 \in \mathbb{N}$ and the subsets $Y_{r,d}^{\text{good}} \subseteq Y$ for $d > r \geq r_0$ be as provided by Proposition 8.18.
- Now let

$$\varepsilon' := \varepsilon/(2/\gamma + 8\eta/\psi),$$

and apply Lemma 9.7 to obtain some $r \geq r_0$ such that for any $d > r$, if

$$I \in \text{Int}(\mathbb{R}), \quad \mathcal{L}^1(I) \leq 4\ell N_d \quad \text{and} \quad m := \left\lfloor \frac{M \mathcal{L}^1(I)}{2^{d-r} N_r^{1/3}} \right\rfloor,$$

then

$$\text{cov}\left(\left((\text{DCM}_{d-r, 4\ell(N_d \geq \dots \geq N_{r+1})}(I))^m, d_{\text{DCM}, m}\right), \delta'''/2\right) \leq \exp(\varepsilon' \mathcal{L}^1(I)).$$

Fix this r , let $Y_d := Y_{r,d}^{\text{good}}$ for all $d > r$, and, for completeness, let $Y_d := \emptyset$ for $d \leq r$. The conclusion of Proposition 8.18 gives $\nu(Y_d) > 1 - \beta$ for all sufficiently large d .

- Finally, having found this r , in addition to the other parameters chosen above, let $\delta > 0$ be given by Proposition 9.4.

Now assume that $d > r$ is sufficiently large and that B , K and J are as in the statement of the lemma. Observe that

$$J' := J + [-\eta\sqrt{N_d}, \eta\sqrt{N_d}] \subseteq B_\ell(\sigma_{[0; N_d]}^y) + [-2\eta\sqrt{N_d}, 2\eta\sqrt{N_d}],$$

so one also has $\mathcal{L}^1(J') \leq 4\ell N_d$ once d is sufficiently large. Let

$$m := \left\lfloor \frac{M \mathcal{L}^1(J')}{2^{d-r} N_r^{1/3}} \right\rfloor.$$

Step 2: The Hamming-like metric. The next step is to prove an analog of the desired bound with $\pi_X^* d_K^{\mathbf{X}}$ in place of $\pi_X^* d_K^{\mathbf{X}, \infty}$ and with diameter δ'' in place of radius δ' .

Given our assumption on U and choice of δ , Proposition 9.4 asserts that

$$U \cap B_{\delta N_d}^{d_{[0; N_d]}^{\mathbf{Y} \times \sigma \mathbf{X}}}(C) \subseteq \{(y, x) \mid \exists (y', x') \in C \text{ and } \mathcal{F} \text{ such that } |\mathcal{F}| = m \\ \text{and } J', x, x', \mathcal{F} \text{ satisfy (P1)–(P3)}\}.$$

Next, for any $d > r$, Lemma 9.7 gives a Borel partition \mathcal{Q} of $\text{DCM}_{d-r, 4\ell(N_d \geq \dots \geq N_{r+1})}(J')^m$ into cells of diameter at most δ''' according to $d_{\text{DCM}, m}$, and with

$$\begin{aligned} |\mathcal{Q}| &\leq \exp(\varepsilon' \mathcal{L}^1(J')) \leq \exp(\varepsilon'(\mathcal{L}^1(J) + 4\eta\sqrt{N_d})) \\ &\leq \exp(\varepsilon'(1/\gamma + 4\eta/\psi)\mathcal{L}^1(K)) \leq \exp(\varepsilon\mathcal{L}^1(K)/2). \end{aligned}$$

The above containment may now be written

$$U \cap B_{\delta N_d}^{d_{[0; N_d]}^{\mathbf{X} \times \sigma \mathbf{X}}}(C) \subseteq \bigcup_{Q \in \mathcal{Q}} R_Q$$

with

$$R_Q := \{(y, x) \mid \exists (y', x') \in C \text{ and } \mathcal{F} \in Q \text{ such that } J', x, x', \mathcal{F} \text{ satisfy (P1)–(P3)}\}.$$

Now observe also that since $\mathcal{L}^1(J) \geq \mathcal{L}^1(K) \geq (2/\psi)\sqrt{N_d}$, we have

$$\mathcal{L}^1(K) \geq \gamma \mathcal{L}^1(J) \geq \gamma \frac{\mathcal{L}^1(J')}{1 + 4\eta/\psi} \geq \gamma \frac{\mathcal{L}^1(J')}{1 + 4/\psi} \geq \gamma' \mathcal{L}^1(J').$$

Therefore, the choice of η and δ''' using Lemma 9.8 implies that

$$\text{diam}(R_Q, \pi_X^* d_K^{\mathbf{X}}) < \zeta \mathcal{L}^1(K) + \zeta \sqrt{N_d} \leq \zeta(1 + 2/\psi) \mathcal{L}^1(K) < \delta'' \mathcal{L}^1(K) \quad \forall Q \in \mathcal{Q}.$$

Step 3: The Bowen-Dinaburg metric It remains to improve our conclusion from $\pi_X^* d_K^{\mathbf{X}}$ to $\pi_X^* d_K^{\mathbf{X}, \infty}$. This follows because, by Lemma 3.7 and our choice of δ'' , each of the sets R_Q obtained above may in turn be covered by at most $\exp(\varepsilon \mathcal{L}^1(K)/2)$ balls of radius δ' for the pseudometric $\pi_X^* d_K^{\mathbf{X}, \infty}$. \square

Proof of Proposition 9.1. This follows from two back-to-back appeals to Lemma 9.10, with some care over the values of all the error tolerances.

Step 1: Choosing the parameters. We are given $\gamma, \psi, \varepsilon, \beta, \delta' > 0$.

By the first part of Lemma 9.10, we may choose some $\eta_1, \delta_1 > 0$ and subsets $Y_{d,1} \subseteq Y$ such that $\nu(Y_{d,1}) > 1 - \beta/2$ for all sufficiently large d , and such that the following holds. If $B' \in C_0(0, 1]$, $J := \sqrt{N_d} B'_{[0,1]}$, and $K \in \text{Int}(J)$ with both

$$\mathcal{L}^1(K) \geq (\psi/2)\sqrt{N_d} \quad \text{and} \quad \mathcal{L}^1(K) \geq \gamma \mathcal{L}^1(J),$$

and if

$$U \subseteq \{y \in Y_{d,1} \mid \|\text{traj}_{N_d}(\sigma^y) - B'\|_\infty < \eta_1\} \times X,$$

then

$$\text{cov}\left((U \cap B_{\delta_1 N_d}^{d_{[0; N_d]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}}(C), \pi_X^* d_K^{\mathbf{X}, \infty}), \delta'\right) \leq \exp(\varepsilon \mathcal{L}^1(K)/2)$$

whenever $C \subseteq U$ has diameter at most δ_1 according to the pseudometric $\pi_X^* d_K^{\mathbf{X}, \infty}$.

Having done so, now the second part of Lemma 9.10 gives some $\eta \in (0, \eta_1]$, $\delta \in (0, \delta_1]$ and subsets $Y_d \subseteq Y_{d,1}$ such that $\nu(Y_d) > 1 - \beta$ for all sufficiently large d , and such that the following holds. If $B \in C_0(0, 1]$, $I := \sqrt{N_d} B_{[0,1]}$, and $K \in \text{Int}(I)$ with both

$$\mathcal{L}^1(K) \geq (\psi/2)\sqrt{N_d} \quad \text{and} \quad \mathcal{L}^1(K) \geq \gamma \mathcal{L}^1(I),$$

and if

$$U' \subseteq \{y \in Y_d \mid \|\text{traj}_{-N_d}(\sigma^y) - B\|_{\infty} < \eta\} \times X,$$

then

$$\text{cov}\left((U' \cap B_{\delta N_d}^{d_{[-N_d; 0]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}}(C), \pi_X^* d_K^{\mathbf{X}, \infty}), \delta_1/2\right) \leq \exp(\varepsilon \mathcal{L}^1(K)/2)$$

whenever $C \subseteq U$ has diameter at most δ according to the pseudometric $\pi_X^* d_K^{\mathbf{X}, \infty}$.

This gives our choice of η , δ and Y_d .

Step 2: Completion of the proof. Now suppose that $B, B' \in C_0(0, 1]$ have $\text{aspect}(B, B') > \gamma$ and $\mathcal{L}^1(B_{[0,1]} \cap B'_{[0,1]}) \geq \psi/2$, and let

$$I := \sqrt{N_d} B_{[0,1]}, \quad J := \sqrt{N_d} B'_{[0,1]}, \quad \text{and} \quad K := I \cap J.$$

Then $\mathcal{L}^1(K) \geq (\psi/2)\sqrt{N_d}$, and the lower bound on $\text{aspect}(B, B')$ implies that $\mathcal{L}^1(I), \mathcal{L}^1(J) \geq \gamma \mathcal{L}^1(K)$. Also set

$$U := \{y \in Y_d \mid \|\text{traj}_{-N_d}(\sigma^y) - B\|_{\infty} < \eta \text{ and } \|\text{traj}_{N_d}(\sigma^y) - B'\|_{\infty} < \eta\} \times X,$$

and suppose that $(y, x) \in U$.

Since

$$U \subseteq \{y \in Y_d \mid \|\text{traj}_{-N_d}(\sigma^y) - B'\|_{\infty} < \eta\} \times X$$

the choice of δ (applied with $C := \{(y, x)\}$) implies that the set $U \cap B_{\delta N_d}^{d_{[-N_d; 0]}^{\mathbf{Y} \times_{\sigma} \mathbf{X}}}(y, x)$ has a Borel partition \mathcal{R} into cells of diameter at most δ_1 according to the pseudometric $\pi_X^* d_K^{\mathbf{X}, \infty}$ and with

$$|\mathcal{R}| \leq \exp(\varepsilon \mathcal{L}^1(I \cap J)/2).$$

Next, for each $C \in \mathcal{R}$, since

$$C \subseteq U \subseteq \{y \in Y_d \mid \|\text{traj}_{N_d}(\sigma^y) - B\|_{\infty} < \eta\} \times X,$$

and since $\delta \leq \delta_1$ and $\eta \leq \eta_1$, the choice of δ_1 and η gives that

$$\text{cov}\left((U \cap B_{\delta N_d}^{d_{[0;N_d]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}}(C), \pi_X^* d_K^{\mathbf{X}, \infty}), \delta'\right) < \exp(\varepsilon \mathcal{L}^1(I \cap J)/2).$$

Since

$$U \cap B_{\delta N_d}^{d_{[0;N_d]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}}(U \cap B_{\delta N_d}^{d_{[-N_d;0]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}}(y, x)) = \bigcup_{C \in \mathcal{R}} (U \cap B_{\delta N_d}^{d_{[0;N_d]}^{\mathbf{Y} \ltimes_{\sigma} \mathbf{X}}}(C)),$$

these bounds combine to give an overall δ' -covering number of the whole bi-neighbourhood, according to $\pi_X^* d_K^{\mathbf{X}, \infty}$, of at most $\exp(\varepsilon \mathcal{L}^1(I \cap J))$. \square

10 Further questions and directions

10.1 Further understanding of the marginal m.p. spaces

Another natural approach to Theorem A would seek an enhancement of Kalikow's proof of Theorem 4.1 which somehow quantifies the failure of either the Very Weak Bernoulli condition or extremality.

An interesting proposal towards this end has been widely discussed by Thouvenot, often in connection with his Weak Pinsker Conjecture. For a general shift-invariant process $(A^{\mathbb{Z}}, \mu, S)$ with marginal m.p. spaces $(A^N, d_{\text{Ham}}, \mu_N)$, he suggests considering the smallest number of pairwise-disjoint subsets of A^N that one needs in order that their union carry most of μ_N , and so that the conditional measure of μ_N on each of them exhibits exponential concentration. We will not define this more carefully here, but refer to it as the 'concentrating-decomposition rate'.

This is an attractive idea in the context of RWRS_{μ} , because the decomposition (11) can be associated with the family of graphs

$$\{(y, F_c(y)) \mid y \in Y_N\} \quad \text{for } c \in X_N,$$

where the notation is as in Section 4. With a little trimming, this decomposition can be turned into a pairwise-disjoint family of subsets of $\{\pm 1\}^N \times C^N$ that carry most of ρ_N , and number roughly $\exp(2R\text{h}(\mu, S)\sqrt{N})$. If one could show that this decomposition is, up to order $\exp(o(\sqrt{N}))$, among the most efficient ways to break ρ_N into exponentially-concentrated components, then it seems that the scenery entropy $\text{h}(\mu, S)$ naturally appears inside this intrinsic geometric invariant of the spaces $(\{\pm 1\}^N \times C^N, d_{\text{Ham}}, \rho_N)$.

Unfortunately, it is not clear that the conditional measures $\rho_{N,c}$ are exponentially concentrated. By definition, we had

$$\rho_{N,c} = (\text{id}, F_c)_* \nu_{1/2}^{\otimes N}, \tag{39}$$

but we have no guarantee that the functions F_c enjoy any ‘approximate continuity’: indeed, it is easy to see that they do not, by slightly modifying Example 4.4.

Thus, there is no reason why the pushforward in (39) should preserve the exponential concentration property of $\nu_{1/2}^{\otimes N}$, and I do not see any other reason why that property should hold for $\rho_{N,c}$. It could be that, in order to decompose ρ_N into exponentially-concentrated measures, one needs to decompose each $\rho_{N,c}$ further by conditioning on some additional properties of a random walk path y , and I know of no very good estimate on the number of further cells that one would need. For the above idea, it would be essential that this further partition for each $\rho_{N,c}$ use at most $\exp(o(\sqrt{N}))$ cells, so that it does not change the leading-order estimate given by the decomposition according to the graphs of F_c .

10.2 Other random walks

Several variants of the RWRS processes do not fall into the class considered by Theorem A.

Perhaps the nearest relatives are those in which the underlying random walk is p -stable for some $p \in (1, 2)$, so that one has an invariance principle for convergence to a p -stable Lévy process. In this case, I suspect that the proofs above can be easily adapted to give the following.

Conjecture 10.1. *If (Y, σ) are the system and cocycle corresponding to a p -stable random walk on \mathbb{Z} for some $p \in (1, 2)$, and if X is a Bernoulli flow, then*

$$\sup_{\kappa > \kappa' > 0} \sup_{\delta > 0} \limsup_{N \rightarrow \infty} \frac{\log \text{BIPACK}_{\alpha, \kappa, \kappa', \delta}(Y \times X, d_{[-N; 0]}^{Y \ltimes_{\sigma} Y}, d_{[0; N]}^{Y \ltimes_{\sigma} X}, \nu \otimes \mu)}{N^{1/p}} = \psi_{p\text{-stab}}(\alpha) h(X),$$

where $\psi_{p\text{-stab}}$ is the obvious analog of ψ_{BM} for the p -stable Lévy process.

On the other hand, the generalization to random walks in \mathbb{Z}^2 , as in [dHS97], is quite different. The problem there is that a typical pair of trajectories $\sigma_{[0; N]}^y, \sigma_{[0; N]}^{y'}$ spend only $o(N)$ amount of time at locations which are visited by both of them. I suspect this implies that *no* information is robustly remembered by both the N -step past and the N -step future, in the sense of the following.

Conjecture 10.2. *Let e_1, e_2 be the usual basis of \mathbb{Z}^2 . Let $Y = (\{\pm e_1, \pm e_2\}^{\mathbb{Z}}, \nu, S)$ be a Bernoulli shift with $\nu = \nu_{(1/4, 1/4, 1/4, 1/4)}^{\otimes \mathbb{Z}}$, let $\sigma : Y \rightarrow \{\pm e_1, \pm e_2\}$ be the time-zero coordinate, and let X be a finite-entropy Bernoulli \mathbb{Z}^2 -system. Then*

$$\text{BIPACK}_{\alpha, \kappa, \kappa', \delta}(Y \times X, d_{[-N; 0]}^{Y \ltimes_{\sigma} X}, d_{[0; N]}^{Y \ltimes_{\sigma} X}, \nu \otimes \mu) = 1$$

for all sufficiently large N , for all $\alpha > 1$, $\kappa > \kappa' > 0$ and $\delta > 0$.

Similar remarks might apply to a p -stable walk if $p < 1$, in which case the occupation measures are no longer absolutely continuous. This conjecture promises the same behaviour for these systems as for Bernoulli systems (Proposition 5.18), even though they are among those shown to be non-Bernoulli by den Hollander and Steif in [dHS97], using an adaptation of Kalikow’s argument. It seems that a different invariant is needed to distinguish these examples one from another.

References

- [Aar97] Jon Aaronson. *An introduction to infinite ergodic theory*, volume 50 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1997.
- [Aar12] Jon Aaronson. Relative complexity of random walks in random sceneries. *Ann. Probab.*, 40(6):2460–2482, 2012.
- [AD01] Jon Aaronson and Manfred Denker. Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps. *Stoch. Dyn.*, 1(2):193–237, 2001.
- [AS72] R. L. Adler and P. C. Shields. Skew products of Bernoulli shifts with rotations. *Israel J. Math.*, 12:215–222, 1972.
- [AS74] R. L. Adler and P. C. Shields. Skew products of Bernoulli shifts with rotations. II. *Israel J. Math.*, 19:228–236, 1974.
- [Bal00] Viviane Baladi. *Positive transfer operators and decay of correlations*, volume 16 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co., Inc., River Edge, NJ, 2000.
- [Bal03] Karen Ball. Entropy and σ -algebra equivalence of certain random walks on random sceneries. *Israel J. Math.*, 137:35–60, 2003.
- [Bil68] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [BK14] Michael Bromberg and Zemer Kosloff. Weak Invariance Principle for the Local Times of Partial Sums of Markov Chains. *J. Theoret. Probab.*, 27(2):493–517, 2014.
- [Bor81] A. N. Borodin. The asymptotic behavior of local times of recurrent random walks with finite variance. *Teor. Veroyatnost. i Primenen.*, 26(4):769–783, 1981.

- [Bow08] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.
- [BS81] L. A. Bunimovich and Ya. G. Sinaĭ. Statistical properties of Lorentz gas with periodic configuration of scatterers. *Comm. Math. Phys.*, 78(4):479–497, 1980/81.
- [CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.
- [dHS97] Frank den Hollander and Jeffrey E. Steif. Mixing properties of the generalized T, T^{-1} -process. *J. Anal. Math.*, 72:165–202, 1997.
- [dHS06] Frank den Hollander and Jeffrey E. Steif. Random walk in random scenery: a survey of some recent results. In *Dynamics & stochasticity*, volume 48 of *IMS Lecture Notes Monogr. Ser.*, pages 53–65. Inst. Math. Statist., Beachwood, OH, 2006.
- [DSV08] Dmitry Dolgopyat, Domokos Szász, and Tamás Varjú. Recurrence properties of planar Lorentz process. *Duke Math. J.*, 142(2):241–281, 2008.
- [Fel76] J. Feldman. New K -automorphisms and a problem of Kakutani. *Israel J. Math.*, 24(1):16–38, 1976.
- [Fel80] Jacob Feldman. r -entropy, equipartition, and Ornstein’s isomorphism theorem in \mathbb{R}^n . *Israel J. Math.*, 36(3-4):321–345, 1980.
- [GH88] Y. Guivarc’h and J. Hardy. Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d’Anosov. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(1):73–98, 1988.
- [Gou05] Sébastien Gouëzel. Berry-Esseen theorem and local limit theorem for non uniformly expanding maps. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(6):997–1024, 2005.
- [Gro01] M. Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, Boston, 2nd edition, 2001.
- [HH01] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001.

- [HHR00] Deborah Hecklen, Christopher Hoffman, and Daniel J. Rudolph. Entropy and dyadic equivalence of random walks on a random scenery. *Adv. Math.*, 156(2):157–179, 2000.
- [Kah85] Jean-Pierre Kahane. *Some Random Series of Functions*. Cambridge University Press, Cambridge, 2nd edition, 1985.
- [Kal82] Steven Arthur Kalikow. T , T^{-1} transformation is not loosely Bernoulli. *Ann. of Math. (2)*, 115(2):393–409, 1982.
- [Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
- [Kat80a] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, (51):137–173, 1980.
- [Kat80b] A. Katok. Smooth non-Bernoulli K -automorphisms. *Invent. Math.*, 61(3):291–299, 1980.
- [Kes98] Harry Kesten. Distinguishing and reconstructing sceneries from observations along random walk paths. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 75–83. Amer. Math. Soc., Providence, RI, 1998.
- [KM10] Steven Kalikow and Randall McCutcheon. *An outline of ergodic theory*, volume 122 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [Led01] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [Lin99] Elon Lindenstrauss. Indistinguishable sceneries. *Random Structures Algorithms*, 14(1):71–86, 1999.
- [Mei74] Isaac Meilijson. Mixing properties of a class of skew-products. *Israel J. Math.*, 19:266–270, 1974.
- [ML] Heinrich Matzinger and Jüri Lember. Scenery reconstruction: an overview. Notes available online at <http://people.math.gatech.edu/~matzi/overview.pdf>.

- [MR03] Heinrich Matzinger and Silke W. W. Rolles. Reconstructing a piece of scenery with polynomially many observations. *Stochastic Process. Appl.*, 107(2):289–300, 2003.
- [MS94] Katalin Marton and Paul C. Shields. The positive-divergence and blowing-up properties. *Israel J. Math.*, 86:331–348, 1994.
- [Nag57] S. V. Nagaev. Some limit theorems for stationary Markov chains. *Teor. Veroyatnost. i Primenen.*, 2:389–416, 1957.
- [NS12] Péter Nándori and Domokos Szász. Lorentz process with shrinking holes in a wall. *Chaos*, 22(2), 2012. 026115.
- [Orn73] Donald S. Ornstein. An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Advances in Math.*, 10:49–62, 1973.
- [OS73] Donald S. Ornstein and Paul C. Shields. An uncountable family of K -automorphisms. *Advances in Math.*, 10:63–88, 1973.
- [OW87] Donald S. Ornstein and Benjamin Weiss. Entropy and isomorphism theorems for actions of amenable groups. *J. Analyse Math.*, 48:1–141, 1987.
- [PP90] William Parry and Mark Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.
- [PS75] Walter Philipp and William Stout. Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* 2, (issue 2, 161):iv+140, 1975.
- [R⁹⁰] Pál Révész. *Random Walk in Random and Non-Random Environments*. World Scientific, Singapore, 1990.
- [Rat79] Marina Ratner. The Cartesian square of the horocycle flow is not loosely Bernoulli. *Israel J. Math.*, 34(1-2):72–96 (1980), 1979.
- [Rud88] Daniel J. Rudolph. Asymptotically Brownian skew products give non-loosely Bernoulli K -automorphisms. *Invent. Math.*, 91(1):105–128, 1988.
- [Rue04] David Ruelle. *Thermodynamic formalism*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2004. The mathematical structures of equilibrium statistical mechanics.

- [Shi96] Paul Shields. *The Ergodic Theory of Discrete Sample Paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 1996.
- [Sin72] Ja. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.
- [ST79] M. Smorodinsky and J.-P. Thouvenot. Bernoulli factors that span a transformation. *Israel J. Math.*, 32(1):39–43, 1979.
- [SV04] Domokos Szász and Tamás Varjú. Local limit theorem for the Lorentz process and its recurrence in the plane. *Ergodic Theory Dynam. Systems*, 24(1):257–278, 2004.
- [Tho02] Jean-Paul Thouvenot. Entropy, isomorphism and equivalence in ergodic theory. In *Handbook of dynamical systems, Vol. 1A*, pages 205–238. North-Holland, Amsterdam, 2002.
- [Var85] V. S. Varadarajan. *Geometry of quantum theory*. Springer-Verlag, New York, second edition, 1985.
- [Ver94] A. M. Vershik. Theory of decreasing sequences of measurable partitions. *Algebra i Analiz*, 6(4):1–68, 1994.
- [Ver00] A. M. Vershik. Dynamic theory of growth in groups: entropy, boundaries, examples. *Uspekhi Mat. Nauk*, 55(4(334)):59–128, 2000.
- [Ver10] A. M. Vershik. Dynamics of metrics in measure spaces and their asymptotic invariants. *Markov Process. Related Fields*, 16(1):169–184, 2010.
- [VZP13] Anatoly M. Vershik, Pavel B. Zatitskiy, and Fedor V. Petrov. Geometry and dynamics of admissible metrics in measure spaces. *Cent. Eur. J. Math.*, 11(3):379–400, 2013.
- [Wei72] Benjamin Weiss. The isomorphism problem in ergodic theory. *Bull. Amer. Math. Soc.*, 78:668–684, 1972.
- [Xia09] Hong Qiang Xia. Multidimensional local central limit theorem of some non-uniformly hyperbolic systems. *Acta Math. Sin. (Engl. Ser.)*, 25(4):565–580, 2009.
- [You98] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998.